Residential home temperature prediction models that use space conditioning experiments and disparate sources of information

Nils Christian Bausch
Dipl.-Ing. (FH)

The thesis is submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of the University of Portsmouth.

The research described in this Dissertation was partially funded by the University of Portsmouth.

April 2014

A model to predict air temperatures inside a residential home was created, which used local and remote environmental sensor data. Space conditioning experiments were carried out in the residential home and models were created to describe the observed temperature data. Results from laboratory space conditioning experiments were used to create a new experimental prediction model. Static inputs of the experimental model were replaced with dynamic inputs and an improved model was created, capable of general prediction application in the residential home. A novel system for space conditioning control was designed, which applied the improved model for air temperature prediction.

Initial prediction models showed prediction error margins of $\pm 1\,^{\circ}\mathrm{C}$. Residential home and laboratory space conditioning experiments were utilized to create non-linear temperature prediction models capable of application outside of the experimental scope. Improved prediction models were based on the experimental models and showed average error margins of $\pm 0.12\,^{\circ}\mathrm{C}$ compared to observed temperature data. A novel system design was proposed, combining the improved model with a traditional heating control to create a new optimal start-stop heating application for residential homes.

Contents

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

The word count of this thesis is 37493 .

(Nils Bausch)

List of Tables

List of Figures

Glossary

1-wire was a bus system developed by Dallas Semiconductor Inc. for device communications. A distinct feature was the ability to use only two wires for data transfer: ground and data/voltage supply.

6LoWPAN IPv6 over Low power Wireless Personal Area Network.

ANN Artificial Neural Network.

API application programming interface.

ASHRAE American Society of Heating, Refrigerating and Air-Conditioning Engineers.

ASP.NET Active Server Pages .NET.

camel case was also known as 'CamelCase' or medial capitals. It described a compound word or an abbreviation where each element began with a capital letter. Common examples were 'PowerPoint' and 'QuickTime'.

CSV Comma Separated Value.

Git was a distributed revision control system initiated by Linus Torvalds.

HAN Home Area Network.

HTML HyperText Markup Language.

HVAC Heating, Ventilation, and Air Conditioning.

IPv6 Internet Protocol version 6.

ISP Internet Service Provider.

LAN Local Area Network.

LOOCV Leave-One-Out Cross Validation.

MAE Mean Absolute Error.

MASE Mean Absolute Scaled Error.

MPC Model Predictive Control.

MSE Mean Squared Error.

natural ventilation was the process of supplying and removing air through an indoor space without the use of mechanical systems.

OOP Object-oriented programming.

PDF Portable Document Format.

PMV Predicted Mean Vote.

PNG Portable Network Graphics.

PPD Predicted Percentage Dissatisfied.

PWSCG Public Weather Service Customer Group.

QoS Quality of Service.

RAE Relative Absolute Error.

Regular Expression were used in specification and recognition of strings in text. Common usage included particular characters, words or patterns of characters.

RFID Radio-frequency identification.

RMSE Root Mean Square Error.

RSE Relative Squared Error.

RSS-feed RSS (Rich Site Summary) feeds enabled publishers to update website data automatically.

Ruby was a programming language, developed by Yukihiro Matsumoto.

Smart Home was a residential home, equipped with electronic devices that could be remotely controlled and/or programmed for specific purposes, such as lighting, heating, and automation of tasks.

space conditioning was the action of heating and/or cooling a defined space.

SQL Structured Query Language.

TO-92 Transistor Outline Package, Case Style 92.

Transact-SQL was a programming language developed by IBM to interact with relational databases, such as Microsoft SQL Server.

URL Uniform Resource Locator.

VB.NET Visual Basic .NET.

web scraping was a computer software technique of extracting information from websites.

X10 was a protocol for communication between eletronic devices connected through powerline or wireless radio used for home automation.

XML Extensible Markup Language.

Acknowledgements

I hereby thank my supervisors Dr Giles Tewkesbury, Dr David Sanders, and Dr Shalini Ramlall for their highly valued input, guidance and their colleague-ship in the Systems Engineering Research Group.

My thanks go to every member of the research group I met and had the honour and pleasure to work with, had a laugh and spent time together outside of university and research. Thank you Dr Simon Chester, Dr Jorge Bergasa, Dr Ian Stott, Dr Ian Rogers, Gareth Lambert and Maulida Butar-Butar; you made this experience so much more enjoyable.

Many thanks also go to all the PhD research students that I met outside of my own research group and wish them all the best in their life and careers.

My thanks also go to the former undergraduate students Robert Severne, Matthew Gibbons and King Wai Li for assistance in my experiments.

My extended thanks go to the School of Engineering staff: you made sure I would not miss my Annual Appraisals, student record updates as well as the weekly emails on how I could develop and improve my skills :). I am very thankful to all the support I received from every person involved.

My special thanks go to my partner Michelle Lin Braby whose support and input I valued very much. She helped me to finish my ‘book’ and allowed us to move on to a life beyond the ‘research student’.

Thanks also go to Professor Jie Tong for granting me a bursary, which made this research possible.

Dedication

To my family

Chapter 1 Introduction

In this Dissertation, prediction models for inside air temperatures in a vacant residential home are presented. The prediction models were created from environmental data recorded between 2010 and 2012, space conditioning experiments carried out in a residential home, and space conditioning experiments in a laboratory setup.

Optimisation of energy usage and energy efficiency in residential homes was an active research area of Smart Homes and space conditioning at the start of the research. The aims in this research included the reduction of energy cost for home owners, the reduction of CO_2 emissions, and the preservation of inhabitant's thermal comfort. Available modelling methods and commercial products used complex mathematical models and algorithms to optimize heating control applications. The marketed solutions primarily targeted large buildings or required proprietary service subscriptions.

1.1 Motivation

A Smart Home was a home equipped with electronic devices that could be remotely controlled and programmed for specific purposes, such as lighting, heating, and automation of tasks. Energy efficiency and optimisation of residential heating was an ongoing research topic in the area of Smart Homes. Standards and calculation methods were available to address the layout and optimisation of heating appliances inside of residential homes. British and international standards were based on well-established research in space heating and cooling. However, these methods relied on expert knowledge or complex mathematical models, both not accessible to every home owner.

The prediction models described in this Dissertation were not reliant on expert knowledge or computationally expensive, and could be applied in residential heating control. Furthermore, the models should be created from environmental and experimental data collected from local and remote sensors. The documented knowledge in this Dissertation should enable the creation of such a heating control system for a real world application.

1.2 Research objectives

The overall aim of the research was to create a prediction model for inside air temperatures in a vacant residential home that used environmental and experimental data from disparate sources.

Specific objectives to reach the aim were as follows:

- To create a system that collects data from local, remote, and aggregated sensors.
- To create prediction models from historical data.

- To conduct space conditioning experiments in a residential home and create temperature prediction models from the experimental data.
- To conduct and analyse laboratory space conditioning experiments.
- To create a new prediction model that can be used in heating control applications for residential homes.

1.3 Methodology

The research initially focused on creating a system to record and visualise data from local and remote environmental sensors for a residential home. The introduction of sensors for heating detection and vacancy detection enabled the recognition of heated and unheated vacant space in the residential home from April 2011. Periods of unheated vacancy were selected as a data source to create temperature prediction models. The unheated vacancies were also used to validate the prediction models. Furthermore, dynamic influences caused by inhabitants were avoided with the selection of unheated vacancies.

The analysis of the sensor data was carried out with a data mining methodology. In other work (Achterbosch, de Jong, Krist-Spit, van der Meulen & Verberne, 1985; Hudson & Underwood, 1999; Balan, Stan & Lapusan, 2009; Hancu, Stan, Lapusan & Donca, 2010) prior knowledge of physical properties of materials and the structure of a house or the application of complex mathematical models and algorithms was necessary to create prediction models. In this work a data mining approach was chosen to find prediction models within the recorded environmental data.

Initial models were created to establish prediction efficiency and benchmarking methods. An investigation into prediction models revealed that errors were small and within the limits of the sensor accuracy. Furthermore the temperature differences during unheated vacancies were small in comparison to daily temperature differences. This showed, that the prediction results were accurate over the limited temperature ranges encountered in unheated vacancies. A different modelling approach was chosen to create iterative models from daily data; in addition, night time data was used to enlarge the temperature difference.

Space conditioning experiments were conducted at a residential home in October 2011. It was hypothesised, that the residential home's response to a step input could help identify and create better prediction models. The step inputs consisted of continuous heating and continuous cooling of the residential home. Due to time constraints, the observed temperatures inside the residential home did not have time to settle at final temperature values. Experimental models were created to describe the observed data. A new modelling approach was proposed to merge heating and cooling models but it was not further investigated due to time constraints.

Laboratory space conditioning experiments were carried out under the supervision of the author in October 2011 and February 2012. The analytical results helped facilitate the creation of new experimental prediction models and model parameters for the residential home. The new experimental prediction models were applied to temperature predictions in unheated vacant space in the residential home.

The experimental prediction models were applied to night time data, where larger temperature differences were observed. Furthermore, periods of unheated night time were undisturbed by dynamic daytime influences. Improved prediction models were created, where static inputs were replaced with dynamic inputs and the prediction models were validated against night time data. The inclusion of dynamic inputs increased the prediction accuracy in the improved models.

The research work concluded with the design of a heating system incorporating the new improved prediction models. The new heating control could be applied in the area of optimal start-stop, where following questions were answered:

- When does the heating need to be turned off to reach temperature T_X at time t_X?
- When does the heating need to be turned on to reach temperature T_Y at time t_Y?

1.4 Research claims

A broad base of research was completed, knowledge documented, and a foundation platform from which further research and development could be effected was created. Novel work included:

- Creation of
  - novel linear regression models that used environmental historical data recorded from local and remote sensors to predict inside air temperature in a residential home.
  - a new non-linear heating model to predict inside air temperatures.
  - a new non-linear cooling model to predict inside air temperatures.
  - novel linear and non-linear prediction models that included web-scraped weather forecast data.
  - a new method to define a static temperature differences input parameter for inside air and wall temperature prediction models.
  - new methods to define dynamic temperature differences input parameters for inside air temperature prediction models.
  - a revised novel non-linear model to predict inside air temperatures that applied knowledge from laboratory space conditioning experiments to prediction models created from experimental data.
  - new software system to collect data from local and remote sources for the creation of prediction models.
  - a new non-linear approximate heating model for a laboratory space conditioning experiment.
  - a new non-linear approximate cooling model for a laboratory space conditioning experiment.
  - a new iterative linear model to predict inside air temperatures that used accumulated environmental data.

  - a new linear model to predict inside air temperatures with different inputs for different prediction lengths.

- Application of
  - a step-input to a residential heating system to create models as an input for a system that used disparate sources of information to predict temperature in a vacant space.
  - non-linear Kohlrausch exponential functions to create novel prediction models.
- Design of a new residential optimal start-stop heating system that incorporates novel non-linear models to predict inside air temperatures.

A key contribution was the design of a new residential heating control system that applied the novel non-linear prediction models for inside air temperatures based on experimental and historical environmental data.

1.5 Overview of the Dissertation

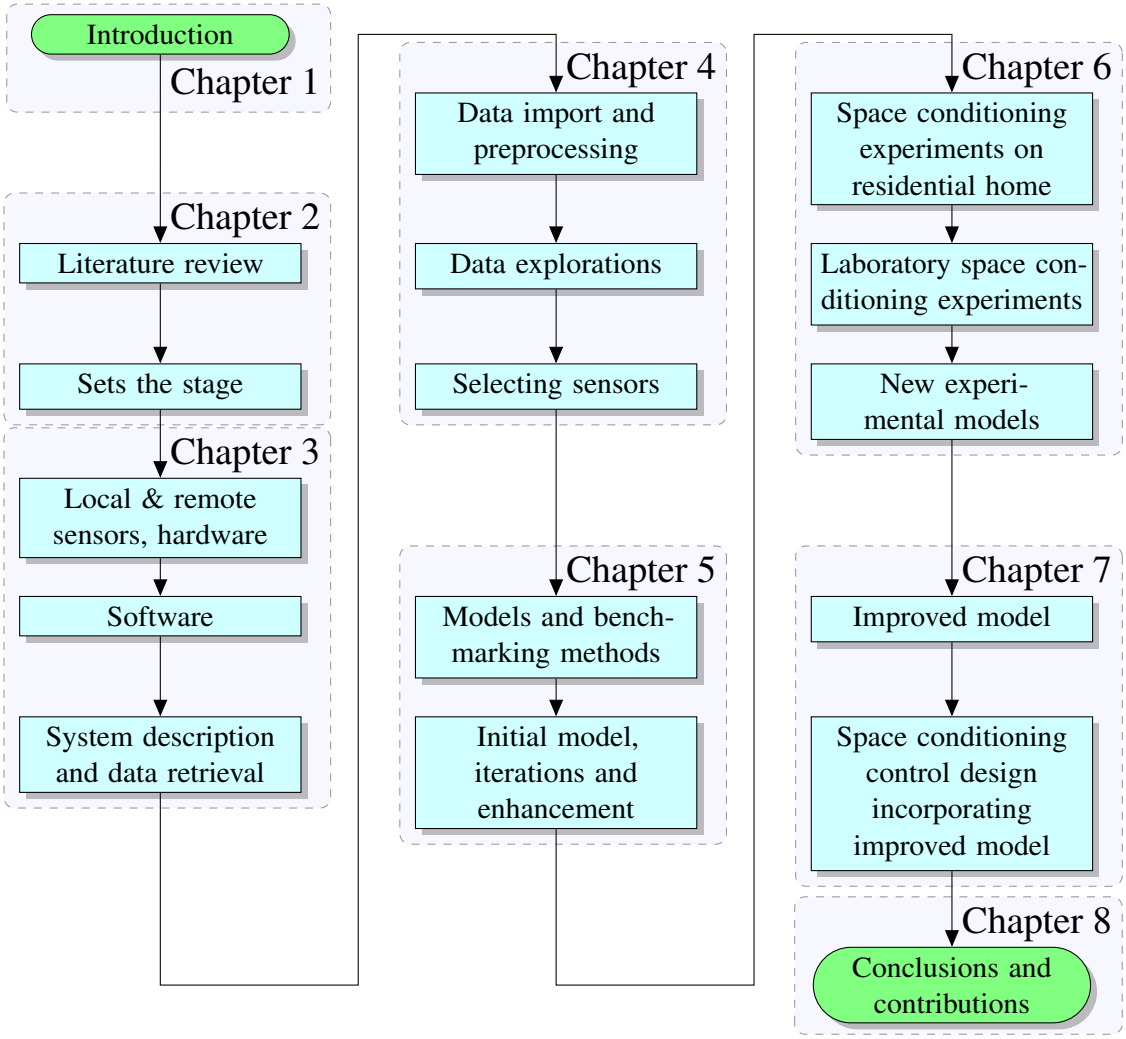


Figure 1.1: *Dissertation overview diagram*

Chapter 2 sets the stage for this research with a review of related literature in the areas of Smart Homes, modelling methods, and space conditioning. The international research is presented and available products discussed.

In Chapter 3 the environmental sensors and the data collection are described. The Chapter introduces the local and remote data sources that were included in the data collection system, which were collectively referred to as disparate sources of information.

Chapter 4 introduces the methodologies of data mining, which were applied to the collected sensor data. The Chapter then continues with an initial data exploration and concludes with an initial selection of sensors for prediction models.

In Chapter 5, methods for creation, comparison, and benchmarking models are presented. The Chapter continues with initial models created from historical environmental sensor data and iterative improvements to the models.

In Chapter 6, space conditioning experiments carried out in the residential home are presented. The Chapter then introduces experimental prediction models, which were created from experimental environmental data. Laboratory space conditioning experiment results are presented and experimental prediction models are introduced at the end of the Chapter.

Chapter 7 presents the creation and application of improved prediction models based on the generalised prediction models from Chapter 6. Chapter 7 concludes with the design of a new heating control application incorporating the novel prediction model.

The Dissertation concludes with Chapter 8, research claims, and recommendations for future work.

A flowchart diagram showing key stages in each Chapter can be seen in fig. 1.1.

Chapter 2 Literature

The aim of this literature review was to introduce current and past research in Smart Homes, space modelling and conditioning. The area of Smart Homes covered multiple subjects, such as computer science, medicine, thermal comfort, entertainment, energy efficiency, optimisation, space heating and modelling. The notion of Smart Homes was mentioned and researched in past literature, either as 'pervasive' or 'ubiquitous computing' (Weiser, 1991). 'Ubiquitous space' was the implementation of computer technology into every day life with as little obstruction as possible, aiding humans in daily tasks at home. Space conditioning was an important factor in Smart Homes to reduce energy costs for the home owners and to drive down CO_2 emissions, as heating in the UK accounted for one third of total CO_2 emissions (The UK Government, 2013b).

The literature review is organized as follows: in Section 2.1 the methods and materials included in this review are described. In Section 2.2, key people and their projects, key journals, and key conferences in the subject area are introduced. The Chapter then presents current research projects. This chapter concludes with a discussion in Section 2.7 on issues that need to be addressed and future challenges.

2.1 Methods and materials

Research papers, articles, and books were acquired as digital copies in Portable Document Format (PDF) files, where possible. The websites used to collect and search the referenced material were:

Web of Knowledge http://webofknowledge.com/

IEEE Xplore http://ieeexplore.ieee.org/

Science Direct http://www.sciencedirect.com/

Google Scholar http://scholar.google.co.uk/

Further articles were retrieved from the articles' references and significant authors were investigated for an in-depth analysis of their projects and research activities.

The search criteria for the literature review were:

- Smart Home research and related areas, such as technologies used, areas of application, sensor types, automation and optimization and applications.
- Building and space conditioning, modelling methods, standards, and applications.

- Reviews about current research in the area of Smart Homes, energy, climate change and modelling.

Keywords identified and used in queries on the aforementioned websites were: 'pervasive space', 'smart home', 'smart environment', 'smart space', 'ubiquitous', 'domotic', 'intelligent environments', 'building automation', 'assisted living', 'sensor fusion', 'prediction', 'lumped components', 'forecast'.

These keywords were used either alone or in conjunction. The keywords were identified through reviewing the literature and added subsequently when new literature was found.

2.2 Research area

The following Section introduces the research area of Smart Homes and building and space conditioning. Key researchers, journals, and conferences that had an impact and influenced the subject area, are included.

2.2.1 Key researchers

The key researchers in the area of Smart Homes and modelling were identified by their contribution to research papers. As the topic of Smart Homes was diverse, researchers tended to focus on several areas.

Lorcan Coyle worked at the University College Dublin in the School of Computer Science and Informatics. His research area was in middleware design, sensor-fusion and pervasive computing. The project 'Construct' was pursued by his research group (Coyle et al., 2007; Coyle, 2009) and it was a software package capable of interconnecting disparate sources of information for sensor-fusion of data.

Chris Nugent was a researcher at the University of Ulster in the School of Computing and Mathematics. The areas covered in his research on Smart Homes were AI algorithms and methods, and sensors. Publications included the influential book 'Designing Smart Homes - The Role of Artificial Intelligence' (Augusto & Nugent, 2006).

Diane Cook was a researcher that worked on AI, sensor data processing, and prediction algorithms for use inside Smart Homes. The published work included the MavHome project (Cook et al., 2003; Das, Cook, Battacharya, Heierman & Lin, 2002) and work on sensor processing (Cook, 2007). She was a lecturer at the Washington State University in the School of Electrical Engineering and Computer Science.

Wolfgang Kastner was a researcher from the Institute of Computer Aided Automation in the Automation Systems Group at the Technical University in Vienna/Austria. His research was in data communication and network technologies used in Smart Home and building automation systems (Kastner, Neugschwandtner, Soucek & Newman, 2005).

Frauke Oldewurtel was a researcher at the Power Systems Laboratory at the ETH in Zürich, Switzerland. Her work included the OptiControl project (Oldewurtel et al., 2012), which

aimed at reducing energy consumption at modest investment and operating costs, while at the same time improving occupant comfort and reducing peak power demand. OptiControl combined developments from the fields of building technologies, numerical weather forecasting and control engineering.

2.2.2 Key journals

Key journals with frequent publication of articles on Smart Home and space heating and modelling research were:

IEEE Pervasive Computing (ISSN 1536-1268) a journal that published articles about ubiquitous and pervasive computing from the IEEE with numerous articles focussing on Smart Homes.

Computer (ISSN 0018-9162) was an IEEE journal and covered the general area of computing and technology with articles about Smart Home developments and also technologies involved in creating them, for example sensor networks, communications, and AI advancements.

Pervasive and Mobile Computing (ISSN 1574-1192) was a journal published by Elsevier that printed articles about advancements in computing. The multiple areas published included security issues, tracking technologies, wireless sensor networks, RFID, autonomic environments, and middleware.

Energy and Buildings (ISSN 0378-7788) was an international journal publishing articles with links to energy use in buildings. The aim was to present new research results, and new proven practice aimed at reducing energy needs of a building and improving indoor environment quality.

Building and Environment (ISSN 0360-1323) was an international journal that published original papers and review articles on research, technology and tool development related to building science and human interaction with the built environment in addition to applications to building design and operation.

2.2.3 Key conferences

Conferences that were found to be relevant in the area were:

International Conference on Smart Homes and Health Telematics was an annual conference that presented work in topic areas such as telehealthcare, Smart Home environments, assistive robotics, and user tracking. The publications were found in the Lecture Notes on Computer Science series published by Springer.

International Conference on Ubiquitous Computing was a conference that reported advances in ubiquitous computing areas including intelligent environments and Smart Homes. The conference was annual and the proceedings were published by the Association for Computing Machinery (ACM).

International Conference on Pervasive Computing was a conference initiated by leaders in industry, for example Panasonic, Hitachi and Microsoft Research.

2.2.4 Key book series

Book series that published proceedings and special subject issues were:

Lecture Notes in Computer Science ; a book series published by Springer. The series covered conference proceedings from the International Conference on Smart Homes and Health Telematics (ICOST) and International Conference on Pervasive Computing in addition to Smart Home subject specific publications.

Assistive Technology Research Series was published by IOS Press, aiming to provide collected edited work, proceedings and workshop summaries of developments in smart environments and assistive technologies for the elderly and disabled.

Ambient Intelligence and Smart Environments was a second book series from IOS Press that began publishing in 2009. The series presented the latest developments in ambient intelligence and smart environments.

2.3 Smart Homes

A Smart Home is defined in this Dissertation as a home equipped with electronic devices that can be remotely controlled and programmed for specific purposes, such as lighting, heating, sensing, and automation of tasks. Research in Smart Homes has been diverse and research activity concentrated not only on fully operational systems but also the technology behind the systems (Chan, Estève, Escriba & Campo, 2008, 2009). The different aspects and building blocks of Smart Homes are presented in this Section, for example: gateways, middleware, methodologies and theories for processing sensor data, communication technology and standards.

This Section is divided into Subsections to introduce research in the different application areas: sensors, used technologies, and commercially available products for Smart Homes.

2.3.1 Application areas

Literature was reviewed in the following areas:

A Ageing at home (Chan, Viard, Caillavet & Campo, 1995; Chapman & McCartney, 2001; Harmo, Taipalus, Knuuttila, Vallet & Halme, 2005; Coyle et al., 2006; Jakkula, Cook & Jain, 2007; Mokhtari, Khalil, Bauchet, Zhang & Nugent, 2009).

B Telehealth and telemedicine (Briggs, Curry & Madge, 2002; Helal, Mitra, Wong, Chang & Mokhtari, 2008; Bonhomme, Campo, Esteve & Guennec, 2007; Zou, Xie & Lin, 2007).

C Energy efficiency optimization and automation (Cook et al., 2003; Helal et al., 2005).

A Ageing at home A study of the Portsmouth Smart Homes project by Chapman and McCartney (2001) identified technology to create accommodation that supported the occupants in achieving an improved independent life; more than would normally be possible because of their physical disabilities. A survey involving disabled people reported positive attitudes to Smart Homes featuring energy efficiency, convenience, security, and safety features. The survey further mentioned that

systems should react to the user's immediate needs, and be able to adapt to changing needs. The respondents wanted a home that could respond to emergencies and environmental changes. Potential occupants also expressed the wish to be able to override automated controls and manually operate all key functions such as doors and windows.

The PROSAFE-extended architecture (Bonhomme et al., 2007) was designed to integrate new services to help the elderly at home, such as: calling emergency doctors, and assisting with medical diagnosis by merging data exploited by intervening services. This architecture used infra-red presence detectors to provide agitation, immobility, location and speed detection functions.

B Telehealth and telemedicine A technology-assisted smart health care system (Jakkula et al., 2007) enabled elderly people to lead an independent lifestyle away from hospitals and avoid having expensive carers. The authors presented a solution where a prediction model in an intelligent smart home system was used to identify health trends over time and enable prediction of future trends which aided in providing preventive measures.

In a review by Chan et al. (2008), the authors investigated Smart Homes, which not only assisted people with reduced physical functions but helped to resolve the social isolation they faced. The authors stated that Smart Homes were capable of providing assistance without limiting or disturbing the resident's daily routine, which increased their comfort, pleasure, and well-being. The findings were that Smart Homes had to deliver a tighter integration with existing residential structure. The proposed solutions had to match or exceed a patient's standard of living to be acceptable. Furthermore user habits and intentions needed to be studied in detail and respected whenever possible.

Based on discussions with medical professionals, patients, educators, and home owners, the authors of the House_n project (Intille, 2002) concluded that the home of most value would be partially automated. The House_n would help inhabitants to learn how to control the environment and the acceptance of automating tasks was increased.

C Energy efficiency and home automation The Gator Tech Smart House (Helal et al., 2005) was a programmable space specifically designed for the elderly and disabled. The project's goal was to create assistive environments such as homes that sensed themselves and their residents and enacted mappings between the physical world and remote monitoring and intervention services. Ultimately, the authors tried to create a 'smart house in a box': off-the-shelf assistive technology for the home that the average user could buy, install, and monitor without the aid of engineers.

The goal of the MavHome project (Das et al., 2002; Cook et al., 2003) was, to create a Smart Home that acted as a rational agent. The agent sought to maximize inhabitant comfort and energy efficiency. To achieve these goals, the agent was able to predict mobility patterns and device usage of the inhabitants. This architecture integrated research in machine learning, databases, mobile computing, robotics, and multimedia computing, which was essential for furthering Smart Home development. This novel approach used movement histories to learn likely future locations of inhabitants and determined which episodes in an inhabitant's history were significant. As a result, these episodes represented events that were going to be automated by MavHome in the future.

The XENO Project by Giorgetti et al. (2008) was a complete solution for home and building

automation . The authors introduced a Home Area Network (HAN) where a digital television service was delivered in the presence of video surveillance, automation, and Internet traffic, sharing the same bandwidth. They described the introduction of a Quality of Service (QoS) router to regulate the priority and bandwidth assigned to services, through the definition of rules. The researchers argued, that the introduction of such an element in the HAN was recommended, in order to avoid bottlenecks and ensure continuous availability of basic and essential home functional services.

2.3.2 Technology

The technology behind Smart Homes made it possible to address issues for ageing at home. The ability to combine different systems and analyse or combine sensor data created a vast number of sensor inputs. Selected research projects that tackled the underlying technology of Smart Homes are addressed in the following paragraphs.

2.3.2.1 Sensors

Integration of sensors into everyday objects, such as tea cups and mobile phones, led to context awareness through sensor fusion, where the location of sensors to each other and their absolute location enabled a context inference (Gellersen, Schmidt & Beigl, 2000).

The Gator Tech Smart House by Helal et al. (2005) incorporated disparate sensors to address different applications: a smart mailbox with notifications for inhabitants; a smart front door with a Radio-frequency identification (RFID) keyless entry; automated smart blinds; a smart bed to monitor sleep patterns; a smart closet to make suggestions based on outdoor weather; smart laundry that helped to sort clothing; a smart bathroom with sensors to detect inhabitants cleanliness or measuring the body weight; a smart microwave that helps with the cooking times and instructions; ultrasonic location tracking to detect inhabitants' movement; location and orientation; a smart floor with embedded pressure sensors to detect location and falls; to name a few.

The PlaceLab home (Logan, Healey, Philipose, Tapia & Intille, 2007) contained over 900 sensor inputs, including wired reed switches, current and water flow inputs, object and person motion detectors, and RFID tags. The authors' aim was to compare different sensor modalities on data that approached 'real world' conditions. They found that 10 infra-red motion detectors outperformed the other sensors on the activities studied, such as 'dish washing' and 'actively watching TV', especially those that were typically performed in the same location.

Ultrasonic sensors were used for monitoring of patients and the elderly in research led by Pham, Qiu, Wai and Biswas (2007). The results showed potential for the usability of ultrasonic sensors in monitoring indoor movements of people, and in capturing and classifying movement trajectories.

Research undertaken by Hong et al. (2009) aimed at a collection of sensing technology which was used to monitor the behaviour of an inhabitant and their interactions with the environment. According to the researchers, information analysis was obstructed by sensors which were not consistently able to provide reliable information due to either faults, operational tolerance levels or corrupted data. In their paper they addressed the fusion process for contextual information derived from uncertain sensor data. Based on a series of information handling techniques evidential contextual information was represented, analysed and merged to achieve a consensus in automatically

inferring activities of daily living for inhabitants in Smart Homes.

With homes that were built with smart technology from the ground up, research led by Hussain, Schaffner and Moseychuck (2009) argued that it would be interesting to see what kind of additional sensors could be used to make a Smart Home. Smart homes that rely heavily on radio signals would have to use construction materials that were less obstructive to radio signals for walls and doors in order to reduce noise. With RFID, Smart Homes could have tags embedded in nearly every item in the house. Furniture, dishes, even clothing could have RFID tags embedded into them. If every object in the house had a tag then the number of possible applications of their system increased dramatically. In their experimental layout, the researchers found, that RFID could help find misplaced items, or let an inhabitant keep track of what their pet was doing during the day.

The self-configuration, self-optimization, self-healing and self-protection of sensors and devices in a HAN were addressed in research by de Vergara et al. (2008). An autonomic element was developed with the goal of finding and personalizing the service offer for the inhabitants. The autonomic element was context aware, as it sensed devices connected to the home network together with user preferences. The services and profiles were modelled and reasoning rules created, which were used by the autonomic element for inferring useful services and proposing a personalized service offer for each inhabitant.

2.3.2.2 Interoperation and communication

Different Smart Home research projects and marketed products in general did not have a common hardware or software interface for data and signal communication exchanges. Therefore research into this area and standards emerged to tackle this problem.

DomoML (Furfari, Sommaruga, Soria & Fresco, 2004) was a mark-up language aimed at the definition of an interoperability standard for domestic resources. The project focused on human language as a means for mediating user interaction with the Smart Home environment, enabling a user to control, query and program devices. Various heterogeneous components took part in this architecture, and DomoML was the glue which interfaced them.

Research led by Nugent et al. (2007) presented the outline design of homeML, an Extensible Markup Language (XML) based schema for representation of information within Smart Homes. This approach provided a common platform for the exchange of data between heterogeneous systems.

The Universal Home Network Middleware architecture (K.-D. Moon, Lee, Son & Kim, 2003) ensured seamless interoperability, and provided scalability by simply adding an Adaptor for each of the Smart Home systems. Moreover, the system provided an environment for the deployment of Smart Home network, and dynamically created new services by combining the functions of appliances without being limited by the individual system.

Other approaches for potential customers to integrate Smart Home or Building Automation Systems into their homes was through open standardised communication protocols, such as KNX, BACnet or LonWorks (Kastner et al., 2005). The open standards (KNX Association, 2013) averted a vendor lock-in for customers and introduced gateways for interoperability with different open standards.

2.3.3 Commercial products

The integration of Smart Home technology was implemented in different ways, dependant on the necessities of the inhabitants. Implementations took into account whether the building was residential or used as an office space and if the technology could be retrofitted or were factored in at the time of construction.

Retrofitted solutions used the power line as communication medium in X10. Wireless technologies were introduced to make inhabitants independent in their choice of placing sensors and actuators. Systems such as WeMo and WigWag (Belkin, 2013; Hemphill, 2013) were internet enabled and integrated with tablet devices and smart phones for remote control or third party web applications such as 'If-This-Then-That' (Tibbets, 2013).

The solution offered by PassivSystems (2013) was built on top of an adaptable wireless system, which would learn within one week from inhabitants' inputs and adjust warm water and indoor heating for summer, winter and transition periods to the inhabitants' needs. The system would override previously learned procedures if an inhabitant repeatedly entered a new desired setting.

The IPv6 over Low power Wireless Personal Area Network (6LoWPAN) standard addressed security and power issues commonly found in standard wireless communication methods. The technology was used by Lietzow, Dalheimer and Walk (2013) in the HeXaBus system, which turned everyday devices into parts of the 'Internet of Things', through the use of Internet Protocol version 6 (IPv6) and low-power consumption hardware. The IPv6 accommodated approximately 3.4×10^{38} device addresses and made every device accessible from the internet, if the Internet Service Provider (ISP) implemented this standard.

A Smart Home and building management protocol called KNX was a standard and also the name of an association which promoted an open system (KNX Association, 2013). The KNX protocol was aimed at new buildings, as it was mainly based on wired connections through twisted pair to accommodate time critical applications, such as emergency and security.

2.3.4 Application to the research

The research described in this Dissertation fitted in the area of Smart Homes, as the intended heating control system described in Section 7.4 on page 124 was targeted at automating and improving a residential heating system. A future scenario, describing the interaction of inhabitants and this new system can be found on Section 7.5 on page 127.

The data collection system described in this Dissertation did not use any of the technologies mentioned in this literature review, as the sensor network was already in place before the research started.

Disparate sources of information, as discussed by Coyle et al. (2006), were used in the data collection system to increase the diversity of information that could be used for data mining.

2.4 Building and space conditioning

The exact modelling of residential homes and building offices could be important to reduce carbon emissions in the future. The UK government wanted to tackle this by reducing carbon emissions from the housing market by 29% by 2020 (Department of Energy and Climate Change, 2010).

For traditional buildings in Ireland (Government of Ireland, 2010), non-intrusive upgrading measures such as draught proofing, attic or loft insulation and boiler replacement were considered and ensured that a traditional building had the potential to out-perform a newly built building over a lifetime of one hundred years.

Climate change and its effects might have an impact on energy prices and bills in the future (The UK Government, 2013a) and recent reports (The UK Government, 2012) showed, that the estimated costs for heating might decrease due to elevated ambient temperatures, which might bring with it health issues in the future if not addressed properly (Bone, Murray, Myers, Dengel & Crump, 2010).

Thermal comfort was an important driver for inhabitants to change their behaviour and adapt their home to the newest standards (Clinch & Healy, 2001).

The methods and solutions involved in modelling, predicting and controlling space heating were important in achieving these goals, which are presented in the following Subsections to give an overview of the recent research.

2.4.1 Modelling

The modelling and controlling of space heating research was active in the following areas:

A Modelling through experiments (Achterbosch et al., 1985) or equivalent electrical networks (Hudson & Underwood, 1999).

B The calculation of the envelope in standards (BSI, 2007a, 2007b, 2008).

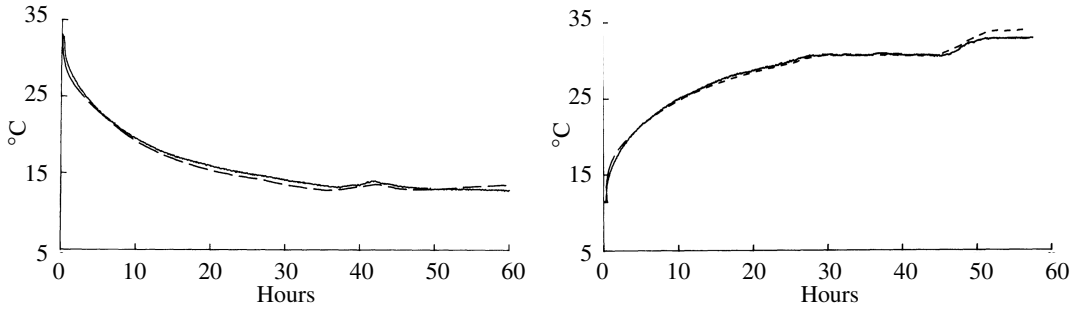
C Taking into account the thermal mass of a building (Balaras, 1996; Kolokotroni, Perera, Azzi & Virk, 2001; Kosny et al., 2001).

D Algorithms, predictions and non-linear models (Grünenfelder, 1985; Chalabi, Bailey & Wilkinson, 1996; Azzi, Loveday, Azad & Virk, 1997; Berglund & Lundberg, 2000; Nielsen & Madsen, 2000; Gegov, 2001; Skrjanc, Zupancic, Furlan & Krainer, 2001; Cho & Zaheeruddin, 2003; Yang & K.-W. Kim, 2004; Ruano, Crispim, Conceição & Lúcio, 2006; Balan et al., 2009; Hancu et al., 2010; J. W. Moon & J.-J. Kim, 2010; Dombayci, 2010; Oldewurtel et al., 2012; Cigler, 2013).

A Experiments and equivalent electrical networks The dynamic behaviour observed in space heating and control was due to higher order systems and researchers tried to formulate and describe these by creating equivalent electrical circuits with lumped components representing different parts involved in heating in a house (Pahwa & Brice, 1985; Achterbosch et al., 1985; Hudson & Underwood, 1999; Parnis, 2012). Resistors would represent walls, capacitors the air inside a room, or a zone and voltages would represent the temperature. Through simplifications and the use of linear equation systems, the higher order systems were simplified to create reduced computationally intensive simulations and design methods for space heating.

Achterbosch et al. (1985) described physical structures of a building and heat inputs with models consisting of electrical components, such as resistors, capacitors and power sources. The researchers created one model representing a single-room approximation and a second model for

a multiple-room setup of a residential home. An experimental part of their research subjected a wooden skeleton and a concrete built house to a step input of heating on and off cycles to test their thermal building models. The thermal building models of the inside air temperatures and recorded temperatures can be seen in fig. 2.1.



(a) Natural cooling *(b) Heating*

Figure 2.1: *Observed and predicted temperatures as solid and dashed lines from experiments conducted by Achterbosch, de Jong, Krist-Spit, van der Meulen and Verberne (1985) on a wooden structure house*

B Standards The British Standards Institute and international institutes developed standards and adapted known calculation methods to create tools for engineers, architects or home owners to estimate their energy efficiency and take optimisation measures (BSI, 2007a, 2007b, 2008, 2012). The standards changed to include newest research carried out in adjacent research areas, such as Thermal Comfort (ASHRAE, 2010; BSI, 2005). The formulas generally used geometrical measurements of the house in question, tabular material constants, and generalised coefficients found by research or as average values.

C Thermal mass The thermal mass of a building was the summation of walls and materials. The materials' characteristics determined the increase and decrease in heat, which could be used to off-set timings for heating and cooling control applications (Balaras, 1996; Kolokotroni et al., 2001; Kosny et al., 2001; Braun, 2003).

Research led by Guerra Santin, Itard and Visscher (2009) showed, that building characteristics contributed to a large part of the energy use in a dwelling (42%) and were therefore not negligible when estimating and controlling space heating systems.

Kolokotroni et al. (2001) found, that non-domestic buildings, which use natural means to provide ventilation can increase indoor air quality and thermal comfort. From their analysis, it was concluded that natural ventilation coupled with thermal mass was able to reduce the effect of external hot weather and establish comfortable conditions within the building. The researchers showed that simple controls based on external temperature, radiation and internal gains, were able to improve the performance of the building, so that periods of uncomfortable conditions were avoided under certain conditions during the summer.

D Non-linear methods Attempts were made with Artificial Neural Networks (ANNs) to determine characteristics, predict temperature development and control space heating (Yang & K.-W. Kim, 2004; Ruano et al., 2006; J. W. Moon & J.-J. Kim, 2010; Bertini et al., 2010; Dombayci,

2010). In general results have shown remarkable improvements in accuracy compared to simplified methods, with the downside, that the ANN description of inputs' relationships to the output was not comprehensible by humans and hidden away in the neural network structure. ANNs had several inputs, which could be current temperature, geographical location, average monthly temperatures, height above sea level, or day and time of the year (Bertini et al., 2010). The output of the ANNs for thermal modelling was a temperature value, which was used to predict temperature values inside a defined space, such as a residential home or a single room. ANNs were a network of interconnected nodes with transition functions and trained with inputs and observed outputs to create a prediction model, when presented with new inputs.

Model Predictive Control (MPC) was a further area intensively researched (Cho & Zaheer-uddin, 2003; Balan et al., 2009; Hancu et al., 2010; Oldewurtel et al., 2010, 2012; Privara, Siroky, Ferkl & Cigler, 2011). The aim of MPC was to control a non-linear system with linear algebra, which could be applied to discrete short operation ranges. A different method involved non-linear models, which could be derived from ANNs. The MPC would control a desired output, such as temperatures, with input variables, such as weather forecasts or local sensor values, and keep values close to a desired range. Results showed potential energy savings of 17–30% compared to traditional controllers Cigler (2013).

Weather forecasts were used in some of the research projects named above. The weather forecast data was used as inputs for prediction models (Grünenfelder, 1985; Privara et al., 2011), which showed improvements. The improvements were dependant on the accuracy of the weather data and improved with better weather forecast models (The Met Office, 2013b).

2.4.2 Commercial products and solutions

The commercially available products enhancing space heating were in general related to or part of Smart Home solutions. The PassivEnergy system (PassivSystems, 2013) was built around wireless sensors to learn inhabitants' preferences on temperatures and to adapt to the inhabitants' vacant times to adjust the heating as needed. Nest (Fadell, 2013) gave users a visually appealing and easy to understand control, which consisted of a big dial and integrated display. Once installed, a learning thermostat replaced the traditional thermostat. The Nest device learned from user inputs and sensors about use patterns, vacancies and temperature changes within a house. In April 2014, Nest brought the learning thermostat to the UK.

The system Termofassade (Schwan, 2010) was not a heating control strategy, but a different method of space heating. The product was a plaster board with an integrated carbon fibre reinforced electric heating mesh, which was installed outside atop the walls of a house. The energy efficiency, thermal comfort and drying of walls were all contributions of the system and were even increased, when solar panels were used to power the system, further decreasing the overall energy consumption.

2.4.3 Application to the research

Thermal comfort was considered a target in Smart Homes and ageing at home at the beginning of the research, but was discarded once the data was analysed. The reasoning was, that preferences collected from inhabitants could be used to create a database, which could be used for a secondary input into the new heating control system. However, the results showed a dissonance in the users'

preferences, which made the collected data unusable in terms of finding a common thermal comfort that was desired by all inhabitants. If a common preference of thermal comfort could be identified, it could be used as an additional goal to the heating control of a residential home. Furthermore the dynamics of the inhabitants' were not recorded with the method described in this Dissertation, such as activity level or clothing. Improvements could be made by increasing the number of respondents and residential homes, to analyse if the dissent found within the researched residential home was a general case or an outlier. A short description of thermal comfort and how it was collected can be found in Appendix C on page 147.

The research described in this Dissertation did not intend to create temperature prediction models with the knowledge of calculations, as outlined by the standards mentioned. Instead the research focussed on using experimental data to infer models. The non-linear methods and algorithms were not considered either, as prior knowledge was used to create these systems.

The most influential work that was applied in parts to the research was from Achterbosch et al. (1985) in that the system response of a residential home was recorded and used to create prediction models.

2.5 Data Mining

Data mining described the process of discovering patterns in datasets, applying methods from statistics, machine learning and artificial intelligence. An important task for preparing data for data mining was the preprocessing. During the preprocessing, numerous steps were applied to the data (Myatt, 2007; Myatt & Johnson, 2009; Kantardzic, 2011). The following list was compiled from books and applied to the data collected in this Dissertation:

1. Importing and collecting the data from all sources.
2. Characterize data and map values of sources if necessary, for example mapping of categorical data to numerical data; 'NE' converted to 90º.
3. Cleaning of erroneous data by deletion or replacement.
4. Remove redundant data.
5. Transform the data to the necessary values needed by the modelling methods.

Upon completion of these preprocessing steps, methods to create and discover models in the data were applied, namely regression analysis, neural networks or correlation analysis. One essential difference of Data Mining compared to traditional modelling methods was that there was no need for prior knowledge of the surveyed system.

2.5.1 Application to the research

Data mining was applied in this research to the data collected from disparate sources of information. This method was chosen, because no prior knowledge was necessary to create models. The description and methods used are described in Chapter 4 on page 36.

2.6 Prediction quality and efficiency measures

Assessment of predictive performance of models was carried out with benchmarking and comparison values. The mathematical measures used in this Dissertation to describe prediction performance are presented in this Section.

2.6.0.1 R^2

A frequently used quality measure in statistics and model predictions was called R^2, also known as coefficient of determination. It represented the model's ability to describe the variation of the predicted data in comparison to the observed data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.1}$$

y_i was the observed value, $\hat{y}_i$ the predicted and $\bar{y}$ the mean value of the observed values, and n was the number of observations. This naming scheme for variables was used throughout the prediction measure equations. The usable output values for R^2 varied between '0' and '1', and were interpreted as follows: '1' to state that the model was a perfect fit and '0' for no fit. Values close to '1' were in general referred to as 'a good fit', which was not quantified further. If R^2 was negative, the result was not interpretable with this method of benchmarking.

2.6.0.2 Residual methods

The following methods relied on residuals and their benchmarking against different denominators.

A MSE, MAE and RMSE The first group was comprised of Mean Squared Error (MSE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). In this group, y was the observed value, $\hat{y}$ was the predicted value, and n was the number of observations.

$$\text{MSE} = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n} \tag{2.2}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{2.3}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{n} \tag{2.4}$$

These methods used the number n of observations to determine the average of an absolute or squared error term, returning positive values. They yielded smaller values for better fitting models, for example a MAE value of 0.01 °C was better than a value of 0.1 °C, when temperature data was used. The returned values were not unit-less, resulting in squared units for the MSE and normal units for the MAE and RMSE. The unit was dependant on the observed data y.

B RSE and RAE The second group was benchmarked against the difference between the observed mean value $\bar{y}$ and the observed value y. They were called Relative Squared Error (RSE) and Relative Absolute Error (RAE) and returned unit-less numbers. The value was used to compare models and a smaller value indicated a better model fitting. For example one model had a RSE value of 0.12 and a second model had a value of 0.10, then the second model's predictions were closer to the

observed values.

$$\text{RSE} = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.5}$$

$$\text{RAE} = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{\sum_{i=1}^{n}|y_i - \bar{y}|} \tag{2.6}$$

C Correlation coefficient r The third measure was previously introduced in Chapter 4 eq. (4.22) on page 51, and is used here to describe a linear relationship between two continuous variables: the correlation coefficient r. The correlation coefficient ranged from '-1' to '+1'. '-1' was interpreted as a negative linear relationship, '+1' as a positive linear relationship and '0' as no linear relationship between two variables. A squared correlation coefficient yielded values in a range between '0' an '1'. Values of and close to '1' indicated that a linear relationship existed and '0' that there was no linear relationship. Furthermore, the observed and predicted values could be plotted in a scatter plot, and the shape and distribution of the resulting graph could help identify linear relationships.

D MASE The fourth measure was the Mean Absolute Scaled Error (MASE) derived by Hyndman and Koehler (2006), Hyndman (2006). The method used a benchmarking measure that was robust and comparable across different prediction models. The benchmark that the MASE used was a 'naïve forecast' fitting error. The naïve forecast used the previous observed value as a prediction for a future value. The MASE values were interpreted as follows: a value equal or greater than 1 was a worse prediction model than the naïve forecast and MASE values lower than 1 was a prediction model better than the naïve forecast.

$$MASE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|}\right| \tag{2.7}$$

The numerator in the MASE related to the fitting error between the observed and the predicted values and it was divided by the average naïve forecast fitting error. $y_{t/i}$ was the observed value, $\hat{y}$ the predicted value and y_{i-1} the preceding observed value. The result of the MASE helped determine if a prediction model was better than a naïve forecast.

2.7 Discussion

This review introduced the current research in Smart Homes, space conditioning, modelling and data mining. The presented research projects were selected to show the plurality of methods, solutions, and limitations.

The work presented in this Dissertation focused on the creation of prediction models for temperatures inside a residential home. To achieve this goal, the traditional methods of creating models for building and space heating were not applied. Instead, the process of data mining was used: collecting data from numerous sources and discovering a new model by analysing the data.

Space conditioning experiments were carried out similar to Achterbosch et al. (1985) in that a residential home was subjected to a heating step-input and cooling. Smart Homes have had a

big impact on how people perceived their homes a few decades ago: a lifeless, inanimate place of living. Through automation and optimization or enabling ageing at home, people came into contact with this new concept of a living pervasive space, which would adjust and learn from them. Although from an energy efficiency point of view, total control would be desirable, but may result in a decreased comfort. Therefore Smart Homes addressed energy efficiency and comfort at the same time. This seemed to conclude the goals of a Smart Home, but still the inhabitants' were not comfortable, if all the decisions were taken out of their hands (Chan et al., 2008; Rashidi, Youngblood, Cook & Das, 2007). Future research had to address inhabitant feedback on multiple levels, to accommodate changes in personal needs and preferences.

Privacy in a smart home could be seen at risk, when sensory elements in every object were incorporated, for example RFID or wearable sensors woven in fabric. As an inhabitant was being subjected to a 24/7 observation, the security of their privacy needed to be addressed using methods such as anonymisation, specific selection, or deletion of collected data.

The static methods of estimating the energy performance (BSI, 2007a, 2007b, 2008) of buildings had deficiencies (Meier, 2010; Schwan, 2013); they did not take into account weather changes, solar radiation and changes in temperatures, in regard to the thermal properties of materials for the envelope. Research into dynamic approaches was necessary to change the calculation method and take season and location into consideration as well, giving inhabitants a better tailored view of their homes and improved energy efficiency regarding space conditioning.

Most of the modelling methods described either needed an intricate knowledge of a building before calculations or predictions could be made or had a complex algorithm to describe the physical relationships between building material and temperatures. Future research would need to find even simpler models, close to MPC to create linear models or use calculation methods, which were easily interpretable and could be used on inexpensive hardware to solve the necessary equations in real time as a continuous output for heating control.

The research work began with the creation of a data collection system to record data from disparate sources. This is described in the next Chapter, Chapter 3.

Chapter 3 Data sources

Data collected for this research used various sensor types and differing underlying technologies. This Chapter introduces hardware, software, data retrieval methods and selected sensors that comprised the system of disparate sources. The Chapter first introduces hardware sensors that were installed in a residential home and software that was used to collect the data. The Chapter continues with environmental sensor data retrieved from websites. The Chapter concludes with the data retrieval and storage of the sensor data.

3.1 Local sensors

Sensors were installed in a residential home in locations such as within and on walls, under the roof, outside, and interfaced two electronic systems: an alarm system and heating control. The sensors and the methods of data access are presented in this Section. The location of the residential home can be seen on a map shown in fig. 3.1.

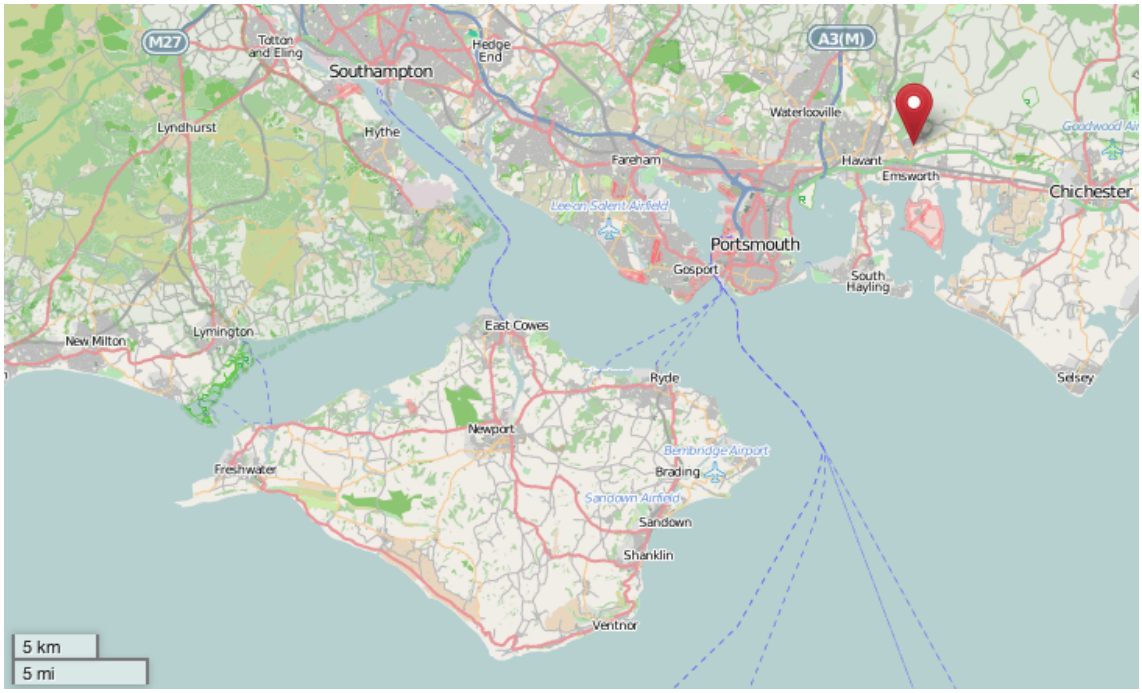


Figure 3.1: *Location of the residential home in the south of England, UK (OpenStreetMap, 2013)*

(a) Single sensor *(b) Sealed sensor with cable* *(c) Interior installation without enclosure* *(d) Exterior wall installation and junction box*

Figure 3.2: *Temperature sensor DS18B20*

3.1.1 1-wire system

The 1-wire[1] system was a communication bus-system developed by Dallas Semiconductor. The 1-wire bus connected devices with two wires: one wire for ground and one wire for data and voltage supply. This was possible due to a built-in capacitor in each device that provided power during data communication. The 1-wire sensor devices in this research work used three wires with a dedicated power supply and a master node to access the network. The use of three wires was necessary due to the number of sensors in the network and the network length to ensure timely data communications. 1-wire sensor devices were generally inexpensive and included for example: digital thermometers, switch sensors, and counters. Each device had a 64 bit long unique identification number, which was used to address devices individually. The following three paragraphs introduce device families that were installed in the residential home in addition to a master device to control the network:

A Digital thermometers The residential home had devices installed prior to the start of the research work in 2009, which were used to collect temperature readings from a swimming pool. These devices were digital thermometers from the device family DS18B20. The digital thermometers had a resolution of up to 12 bit, which enabled a measurement resolution of 0.0625 °C. The Transistor Outline Package, Case Style 92 (TO-92) with its small size enabled an easy deployment of the sensors in walls, plastic casings and outside the residential home. Sensor installations can be seen in fig. 3.2, showing a bare sensor, a sensor sealed in grab adhesive, an inside installation, and an outside installation at the residential home.

B Switch sensors The DS2408, see fig. 3.3, was a switch sensor with eight channels. The channels were user configurable as inputs or outputs. A different two channel switch sensor in use

[1] http://www.maximintegrated.com/products/1-wire/

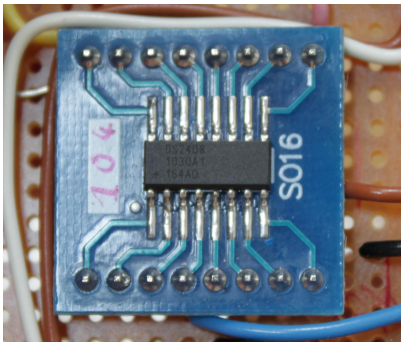


Figure 3.3: *DS2408 on an adapter board*

was from the DS2506 family. The switch sensors were used in interfaces to third party electronic devices.

C HA7net The HA7net was an Ethernet to 1-wire interface, providing a web interface to access 1-wire devices. The HA7net provided HyperText Markup Language (HTML) websites to retrieve and send data to the bus system. Functions that were implemented on the HA7net and applied in this research work were as follows:

- Search: probed the connected 1-wire network and returned all devices or a specific family of devices as a list of unique IDs.
- Write: sent commands to the 1-wire network and returned replies from devices where applicable.
- Temperature read: was a higher level function, which bundled the retrieval of temperature data of single or several devices into a single function.

The HA7net was able to support an overall network length of 300 m with up to 100 connected devices. A picture of the HA7net can be seen in fig. 3.4.

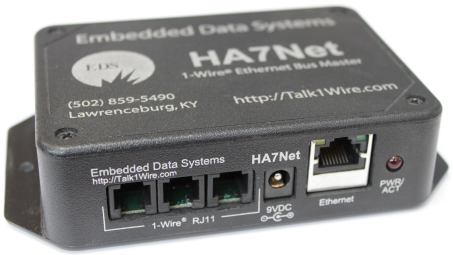


Figure 3.4: *HA7net with three RJ-11 ports for 1-wire and one RJ-45 port for Ethernet*

A diagram showing the HA7net and a 1-wire network can be seen in fig. 3.5.

3.1.2 On-site weather station

Davis Instruments[2] produced a range of weather stations that collected weather data either via wireless or wired communication. The weather station installed at the residential home was a

[2] http://www.davisnet.com/

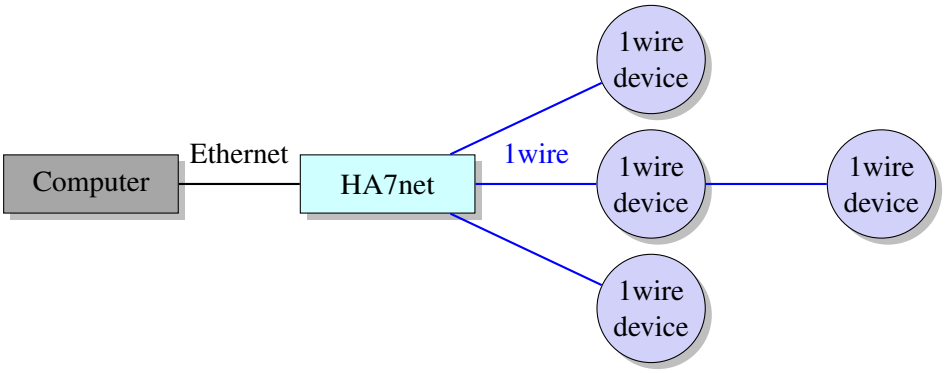


Figure 3.5: *Example network layout of HA7net with connected computer and 1wire devices*

wireless Vantage Pro 2. The weather information was transferred every minute to a console inside the residential home and sent to a website[3] via an Ethernet module. An additional software package enabled the download of historical data from the website. See fig. 3.6 for a weather station and various sensors, such as anemometer, rain collector, wireless transmitter, and console.

Figure 3.6: *Davis Vantage Pro 2 weather station with anemometer, solar panel and wireless transmitter, radiation shield with air temperature sensor, rain collector, and indoor console (Davis Instruments, 2013)*

3.2 Remote sensors

Remote sensors were defined as sensors in the locality of the residential home, which provided environmental data from third party sources. The remote sensors were accessed via their websites, which are described in the following Subsections.

3.2.1 Weather forecast

Websites, such as the Met Office, the BBC and Yahoo provided access to localised weather forecasts as RSS-feed or as a web page. The weather information on these web pages was retrieved through an application programming interface (API) or through matching text from the website to extract data.

The websites provided different environmental weather information, various forecast lengths,

[3] http://www.weatherlink.com

and meteorological details. For example the BBC weather forecast website provided daily forecasts for one, two, and three days. The Met Office website provided daily forecasts of up to five days and hourly forecasts for up to 24 hours.

3.2.2 Weather station

The website for the on-site weather station featured a map and access to the weather station if this was permitted by the weather station owner.

Other local sources of weather conditions were buoys and weather stations collecting data for a nearby harbour in Chichester, Hampshire UK. The buoy websites made available environmental data similar to land based weather stations, in addition to sea conditions.

3.3 Software background

Data collection was carried out with different operating systems, programs and programming languages. Their purpose in the process of data retrieval and their interactions are described here.

3.3.1 Operating systems

Various operating systems were used to access and process the collected sensor data. This was due to the applied programming languages and the setup that was in place before the research work commenced. A Microsoft Windows Server at the residential home collected data from the on-site 1-wire network. The server was accessible from the internet to permit access to a Structured Query Language (SQL) database downloads and visualisations of the recorded data. A Microsoft Windows XP virtual machine was used to import the downloaded SQL database. The websites' data was recorded on a Linux Gentoo server at first, and a secondary system running FreeBSD was set up later on. The final steps of analysis and modelling were carried out on a Mac OS X system in a software package called Matlab.

3.3.2 Programming languages & programs

On-site data collection was achieved with a Visual Basic .NET (VB.NET) program running on the Microsoft Windows Server, named *TempReader*, which saved the data to a SQL database. The program accessed the website of the HA7net device on the Local Area Network (LAN). A preliminary version of this program was available prior to the start of the research in 2009. This version was enhanced by including libraries from the HA7net to implement high level functions. The program was then completely rewritten to support an Object-oriented programming (OOP) design and went through several iterations of improvements and enhancements. An Active Server Pages .NET (ASP.NET) program running on the Microsoft Windows Server was used to create and download backups of the SQL database. An additional ASP.NET program residing on the on-site server, called *Grapher*, created plots and tables with statistics from the collected data. The programming language Ruby was used on a Linux and FreeBSD servers to access the websites, namely weather stations and forecasts. The scripts collected and extracted textual data found on these websites and saved them in separate files for each website.

3.4 System implementation

In this Section the disparate sources of information used in the research work are introduced, consisting of the 1-wire devices, the website forecasts, and the website weather stations. A graph detailing the data retrieval process can be seen in fig. 3.7. The diagram also shows the data import into the software package Matlab, which was used for further data analysis.

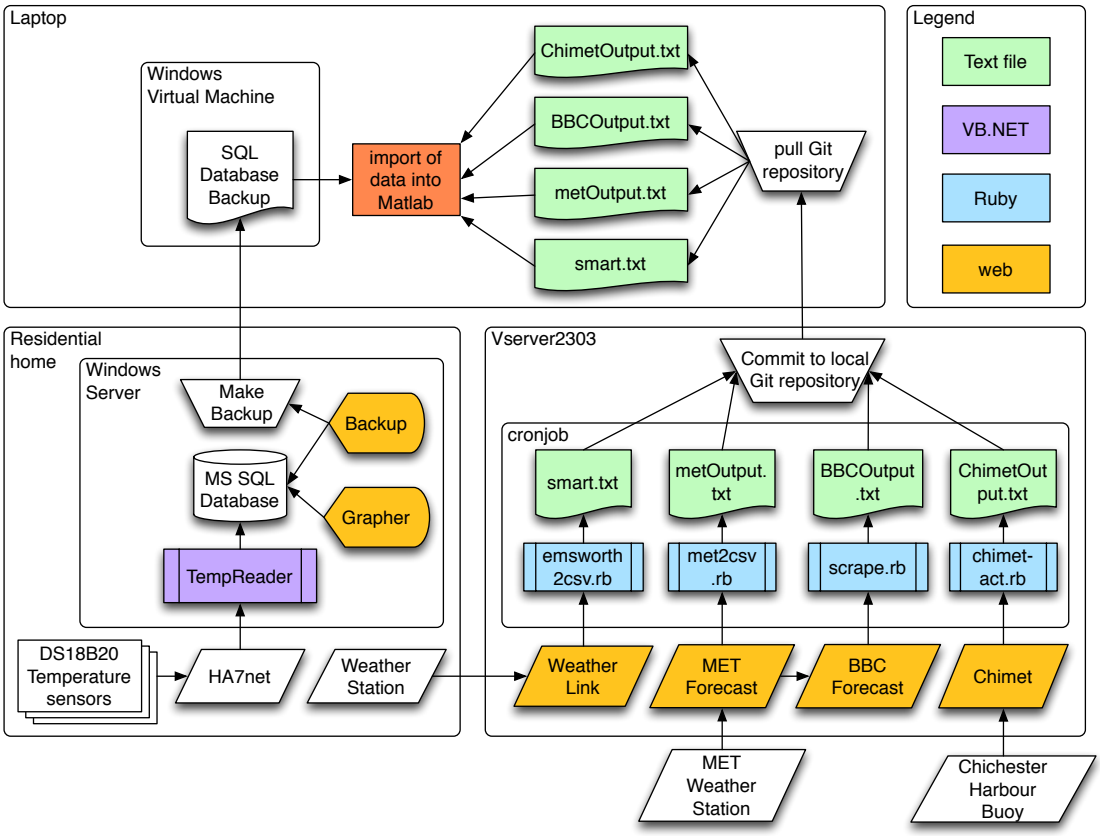


Figure 3.7: *System overview of data sources, showing the sensor data locations in the bottom part and on the top part the data assembly and import into Matlab.*

3.4.1 1-wire devices

A DS2408 switch device was used to interface to a Galaxy 8 house alarm system. The following signals of the house alarm system were interfaced: sounding of the exit horn, alarm unset, alarm set, and intruder status. A schematic of the electrical circuit can be seen in fig. B.3 on page 146 in the Appendix.

A second DS2408 switch device was used in a comfort input box. The comfort input box was used to input the inhabitants' personal thermal comfort level. The supported thermal comfort levels were: cold, ok, and hot. The comfort input box recognized up to five individuals. A schematic for the comfort input box can be seen in fig. B.1 on page 144 in the Appendix.

A third DS2408 switch sensor was used as an interface to the residential home's heating system from Heatmiser. Thermostats in the house signalled a heating demand to the Heatmiser system. By placing a DS2408 onto the Heatmiser control board, the sensor was able to record the heating

demand. A second switch sensor, a DS2406, was placed on the same board. The second sensor was intended to control the Heatmiser heating system by signalling a heating demand, but was not used in this research work due to time constraints. The schematic of the electrical circuit interfacing the Heatmiser system can be seen in fig. B.2 on page 145 in the Appendix. Several DS18B20 temperature sensors were installed inside and outside the residential home. The inside installation comprised of sensors embedded into walls, sensors inside enclosures to collect air temperatures, and sensors exposed to ambient temperatures in different parts of the residential home. The outside temperature sensors were embedded into walls and inside of enclosures to minimize heat radiation affecting the ambient temperature measurements.

3.4.2 On-site weather station

A weather station from Davis Instruments was installed above the apex of the residential home in autumn 2010 to collect local meteorological data. The following sensors values were recorded:

Table 3.1: *Weather station sensors*

Measurement name	Unit
wind speed	$km\,h^{-1}$
wind direction	direction as abbreviated letters / degrees
rain collector	$mm\,h^{-1}$
solar radiation	$W\,m^{-2}$
air temperature	°C
relative humidity	percent
barometric pressure	mbar

The data was transmitted to and displayed on a console inside the residential home. The console submitted the data every minute to a website which made it accessible on http://www.weatherlink.com/user/smart/. The website made available historical data with an interval of 30 minutes, therefore the website with current data was web scraped to collect data in shorter intervals. The WeatherLink website can be seen in fig. 3.8.

Emsworth, Hampshire

Current Conditions as of 10:05 Sunday, March 18, 2012

Station Summary	Current	Today's Highs		Today's Lows	
Outside Temp	8.2 C	8.3 C	09:27	-0.1 C	05:59
Outside Humidity	82%	100%	01:39	82%	09:36
Inside Temp	24.7 C	25.3 C	00:00	21.8 C	07:44
Inside Humidity	27%	30%	00:00	27%	09:08
Heat Index	8.3 C	8.3 C	09:15		
Wind Chill	6.7 C			-0.6 C	04:56
Dew Point	5.6 C	6.7 C	09:06	-1.1 C	05:59
Barometer	1016.6mb	1016.6mb	09:39	1012.3mb	00:00
Bar Trend	Rising Slowly				
Wind Speed	10 km/h	14 km/h	09:22		
Wind Direction	W 268°				
Solar Radiation	617 W/m²	617 W/m²	09:40		
UV Radiation	n/a	0.0 Index	n/a		
12 Hour Forecast	Mostly Clear				

Wind	2 Minute	10 Minute
Average Wind Speed	8.4 km/h	8.0 km/h
Wind Gust Speed		14.5 km/h

Rain	Rate	Day	Storm	Month	Year
Rain	0.0mm/Hour	0.2mm	3.8mm	16.8mm	85.8mm
Last Hour Rain	0.0mm				
ET		0.30mm		24.4mm	77.2mm

Figure 3.8: *WeatherLink website screenshot showing meteorological data transmitted from the on-site weather station*

3.4.3 Weather forecasts

Meteorological forecast data from the BBC[4] and the Met Office[5] were used. The websites provided up to date information on weather forecasts in the area of the residential home.

The BBC website provided a three day forecast, including the current day. Measurements taken from the website are shown in table 3.2.

Table 3.2: *BBC forecast website weather information*

Measurement name	Unit
wind speed	$km\,h^{-1}$
wind direction	direction as abbreviated letters / degrees
air temperature	°C
relative humidity	percent
barometric pressure	mbar
sunrise/sunset	date and time

The Met Office provided a five day forecast, including the current day.The forecasts were updated between two to four times an hour. The weather station site for the Met Office measurements was located approximately 21 km[6] away from the residential home. Measurements that were taken from the website are shown in table 3.3. An example screenshot of the actual website can be seen in fig. 3.9 on the following page.

[4] http://newsrss.bbc.co.uk/weather/forecast/348/Next3DaysRSS.xml?area=Emsworth
[5] http://www.metoffice.gov.uk/weather/uk/se/portsmouth_forecast_weather.html
[6] Calculated with http://www.movable-type.co.uk/scripts/latlong.html

Table 3.3: *Met Office forecast website weather information*

Measurement name	Unit
wind speed	miles per hour
wind direction	directions as letters
air temperature	celsius
visibility	words as scale

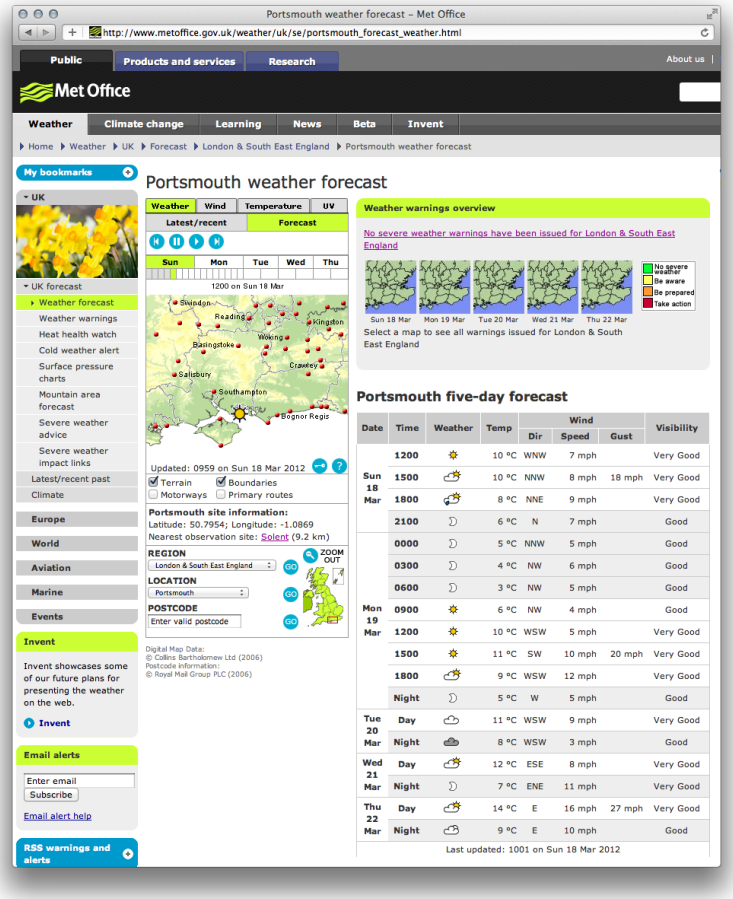


Figure 3.9: *Met Office five day weather forecast website*

3.4.4 Current weather

In addition to the on-site weather station, data from third party weather stations was collected. The first one was installed on a buoy and the second one was a weather station in Portsmouth. The latter was only used to create and test a prototype web scraper for the on-site weather station.

A Chimet The website http://www.chimet.co.uk hosted weather information collected 1.61 km offshore, close to Chichester harbour entrance. The buoy was approximately 11 km away from the residential home. The web scraped readings of the website can be seen in table 3.4

Table 3.4: *Chimet buoy weather information*

Measurement name	Unit
air temperature	°C
barometric pressure	mbar

B Weather station on Portsmouth island A weather station of the same type as described in Subsection 3.4.2 was accessed via its Weatherlink website[7]. The website was used to develop web scraping scripts that were later used for the on-site weather station. The weather station's website frequently stopped updating for days and therefore was not suitable for data collection.

3.5 Retrieval of sensor data

The sensors were categorised by their technologies and differed in their locations in relation to the residential home. The 1-wire network was locally installed and the sensor network data stored on-site. The current weather conditions and forecasts were saved and retrieved on computers located at remote locations. This Section describes the implementation of the data retrieval and storage of the sensor data.

3.5.1 1-wire sensors

The sensors used during this research are listed in table 3.5, which shows the frequency of measurement and the date the sensors became active. The manufacturer of the HA7net created a VB.NET

Table 3.5: *On-site 1-wire sensors with the device family, read interval, and date of first recorded data*

Location	Type	Reading every min	Collecting since
East Wall Outer Sensor	DS18B20	30	2010-07-07
East Wall Inner Sensor	DS18B20	30	2010-07-07
First Floor Core Temp	DS18B20	30	2010-07-07
Silver Tube Air Temp	DS18B20	15	2010-10-13
First Floor Landing Air	DS18B20	15	2010-10-13
Ground East Inner wall	DS18B20	30	2010-11-14
Ground Air	DS18B20	15	2010-11-14
Boiler Cold Water	DS18B20	5	2010-11-24
Boiler Hot Water	DS18B20	5	2010-11-24
Boiler CH Flow	DS18B20	5	2010-11-24
Boiler CH Return	DS18B20	5	2010-11-24
Eaves N-E Air	DS18B20	30	2010-12-29
Roof Under Tiles N-E eaves	DS18B20	30	2010-12-29
Roof Under Tiles N-Mid eaves	DS18B20	30	2011-01-03
Eaves N-Mid Air	DS18B20	30	2011-01-16
Eaves Wall N-Mid Room Side Rad	DS18B20	30	2011-05-13
Eaves Wall N-Mid eaves side	DS18B20	30	2011-05-13
Comfort User 1 Hall	DS2408	10	2011-02-16
Comfort User 2 Hall	DS2408	10	2011-02-18
Comfort User 3 Hall	DS2408	10	2011-02-18
Comfort User 4 Hall	DS2408	10	2011-02-18
Comfort User 5 Hall	DS2408	10	2011-02-18
Alarm Status Sensor	DS2408	5	2011-03-27
Heating Zones	DS2406	5	2011-03-28
Roof Air	DS18B20	30	2011-10-22
Roof Apex	DS18B20	30	2011-10-22
Ground East inner Wall2	DS18B20	30	2011-10-22
Ground Air2	DS18B20	15	2011-10-22

[7] http://www.weatherlink.com/user/geoffs/

class library to access the HA7net's websites and common 1-wire device families. During the development and integration of new sensors, the VB.NET program underwent several changes and additions. The program was able to carry out the following tasks:

1. Retrieve all sensors from the database and initialize them.
2. Check for network connectivity of the HA7net and check each sensors' availability.
3. Initialize a scheduler with the lowest read interval.
4. Scheduled reading of sensors through the HA7net.
5. Saving of the values in a SQL database and logging of errors.
6. Reset schedule timer for next invocation. Repeat from 4.

The program communicated with an SQL server and retrieved initialisation data from the *Sensors* table as shown in fig. 3.10. This table contained following fields: a unique ID used as a reference for connected tables, the unique ROM ID of the sensor, an unused type field, a *RefreshRate* as sensor read interval in minutes used by the scheduler, an unused accuracy value, and a location note as a description.

	ID	Address	Type	RefreshRate	Accuracy	LocationNotes
	1	D2000000C656E428	1	15	0.1000	Environmentally Sealed Sensor
	2	E700000127F64528	1	30	0.1000	Ground Sensor Under Pool (bel...
	3	4C000001286B6128	1	30	0.1000	Floor Sensor Below Pool (abov...
	4	D00000016C749F28	1	15	0.1000	Air Temp Grey Box
	5	BA0000016C3C6528	1	30	0.1000	Ground Center Pool (no Insula...
	6	C80000016C486528	1	30	0.1000	Floor Center Pool (no Insulation)
	7	030000016C6C2628	1	15	0.1000	Pool Water Temp

Figure 3.10: *Screenshot of the sensor table with 1-wire device and setup information*

Data itself was saved in the *TempData* table and used the unique ID from the *Sensors* table, a timestamp, and a value, as can be seen in fig. 3.11. The data was saved by a stored procedure, which was a program written in Transact-SQL and saved inside the SQL database. The stored procedure achieved data compression, by comparing the previous two data values against each other and against the latest value. Depending on timestamps and the values, following outcomes were possible:

1. If the previous two data values had different values, then the latest value would be inserted into the database.
2. If the previous two data values had the same value and were not recorded within one hour then the latest value would be inserted into the database.
3. If the previous two data values had the same value and were recorded within the last hour, the latest value's timestamp would be used to update the previously last value's timestamp. Only in this case the data was compressed, by updating existing data values' timestamp.

An example of this compression, comparing both original collected data and saved data at the end of the process, can be seen in fig. 3.12. The application of compression created data records with irregular read intervals. The original data with regular intervals was recreated with the refresh time during the data import. The time of one hour was chosen to ensure the detection of errors. The highest data compressions were achieved with sensors being read every five minutes, reaching 23%. Sensors with a read interval of 15 minutes reached up to 17% and sensors with a read interval of 30 minutes achieved a compression of 5%.

	SensorID	Time	Value
	1	2009-12-05 14:21:00	38.6875
	1	2009-12-05 14:26:00	38.6875
	1	2009-12-05 14:31:00	38.6250
	1	2009-12-05 14:37:00	38.5625

Figure 3.11: *Screenshot of TempData table with example sensor data, showing columns with an ID related to the sensor table, a timestamp and recorded values*

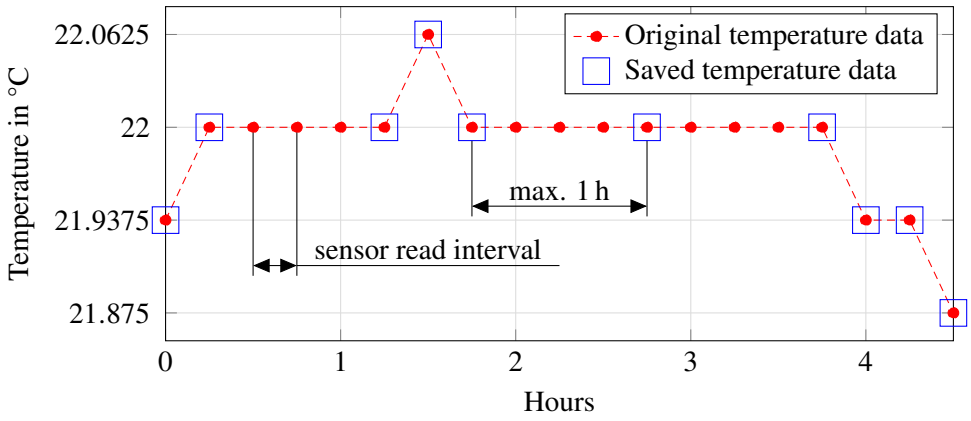


Figure 3.12: *Example data compression showing original and saved data*

Errors from the sensors were logged in an *Errors* table and included the unique ID from *Sensors*, a timestamp, a string error message and an *ErrorID*. A screenshot of this table can be seen in fig. 3.13.

	SensorID	Time	Message	ErrorID
▶	23	2011-07-23 10:54:00	Invalid data received for Eaves N-Mid Air	1
	19	2011-07-24 01:27:00	Invalid data received for Boiler CH Return	1
	17	2011-07-24 19:33:00	Invalid data received for Boiler Hot Water	1

Figure 3.13: *Screenshot of errors table, showing a column for ID, timestamp, error description and an ErrorID*

The SQL server data backup was accessed with an ASP.NET web-interface from http://phd.nilsbausch.net/Backup, which was written by a research group member. A screenshot of the interface can be seen in fig. 3.14 on the following page.

In autumn 2011 a collaborative development with a research group member was started. The output was a graphing tool for historical 1-wire sensor data with a web interface and downloadable

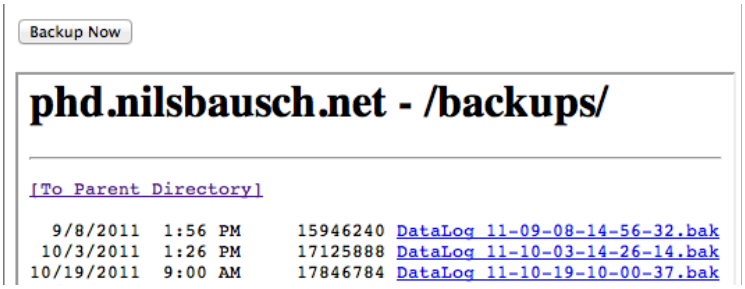


Figure 3.14: *Screenshot of DataLog Backup interface website*

graphs in Portable Network Graphics (PNG) format. The aim was to create a tool, which made available the sensor data and error logs from a remote location. The *Grapher* tool displayed error logs, latest temperatures, and minimum and maximum values of sensors. It was also possible to plot diagrams of different sensors, time spans, and customizations for the output diagrams. An actual graph can be seen in fig. 3.15.

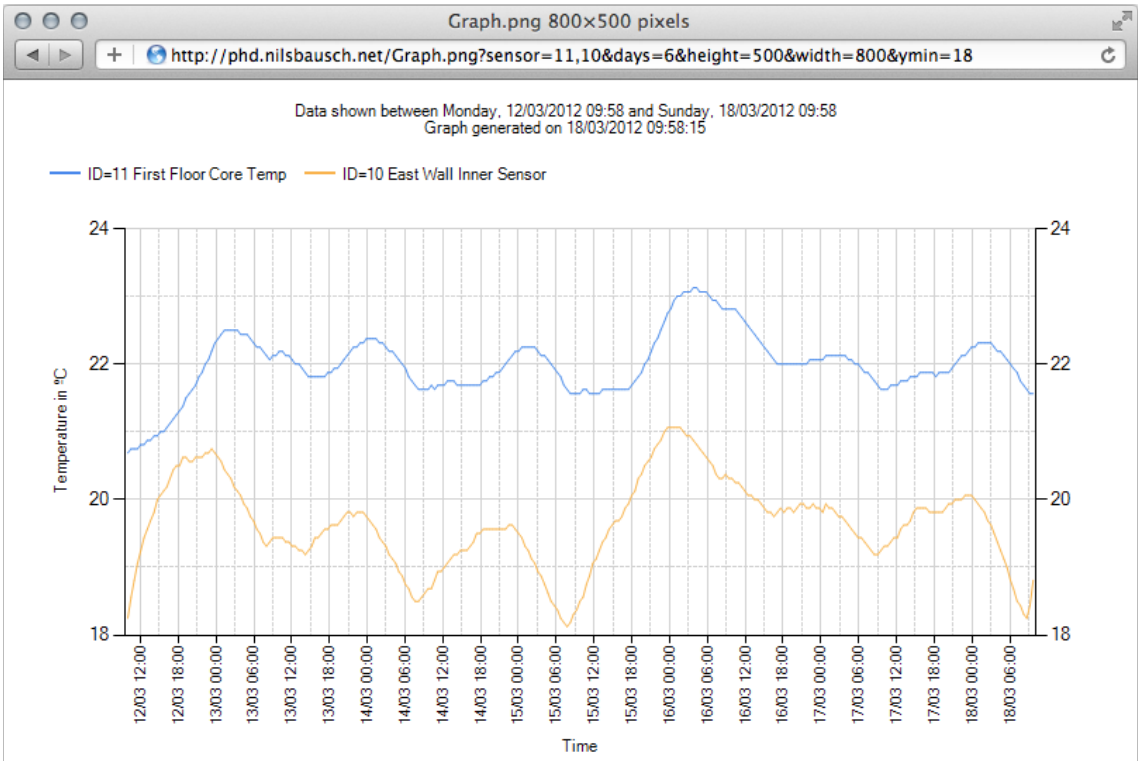


Figure 3.15: *Grapher demo showing PNG output generated by passing arguments to the URL query in the URL bar on top*

3.5.2 Weather forecasts and weather station

The websites were web scraped with Regular Expressions to extract the data, for example temperatures, wind speeds, wind direction or pressure. The web scraping generally progressed in three steps:

1. Access the website via its Uniform Resource Locator (URL).

2. Retrieve data of interest with Regular Expressions.

3. Save the retrieved data and a timestamp in a Comma Separated Value (CSV) file. If the website provided forecasts, a forecast date was recorded as well.

The web scraping scripts were written in Ruby. The scripts were invoked at set intervals to assure that updates on the websites would be detected and saved. The intervals were set as follows:

1. Every minute for the weather station website. The weather station sent updates to the WeatherLink website every minute.

2. Every four minutes for the Chimet buoy website. The website stated, that the data was updated every five minutes.

3. Every five minutes for the Met Office update. The Met Office website was updated between three to five times an hour.

4. Every full hour for the BBC website. The BBC weather feed was updated three to four times a day.

Each of the websites was accessed, the data retrieved, and recorded in separate files. Each file had a headline which listed the data types and units. A headline with example data is shown as follows:

```
publication_date,forecast_date,temp_C,wind_direction,wind_speed_miles
2010-07-21T15:01:00+01:00,2010-07-21T16:00:00+01:00,20,SW,16
```

The filenames for each website were as follows:

Met Office forecast metOutput.txt

BBC forecast BBCOutput.txt

Chimet buoy ChimetOutput.txt

Weather station smart.txt

Davis weather station davis.txt

As the files were saved in a remote location, a backup was made every few weeks with Git to make the data accessible in Matlab for further data manipulation.

3.6 Vacant space and heating sensing

The research described in this Dissertation concentrated on vacant space heating. The model criteria 'vacant space' was introduced in April 2011. The concept behind vacant space was to decrease the dynamics when collecting temperature readings inside the residential home. Presence detection of the inhabitants through an alarm system enabled vacancy detection. Sensors integrated into a boiler collected information on the heating system, enabling heating detection. Both sensors together enabled the detection of vacant space and the heating state of the residential home and were an integral part of the rules set out for defining datasets for the creation of prediction models.

3.7 Conclusion

This Chapter introduced the data collection methods used in this research. Data was accessed and recorded from disparate sources in local and remote locations.

The residential home had sensors installed at the beginning of the research and several additional sensors were installed to collect data of interest, such as temperatures of brickwork, interfacing the alarm system, and inside air temperature. Other data sources were third parties, such as weather station and forecast data, which were accessed through websites. An on-site weather station was used to collect climate data at the residential home. This diverse set of sensors used different technologies and interfaces. Each specific interface to the data sources were presented and described in this Chapter.

The sensors defining vacant space and heating states were introduced in April 2011, marking the start date for data used in later analysis.

The hypothesis was, that diverse sources of information would contribute to the creation of a model which was able to produce usable forecasts.

This Chapter laid the groundwork for data collection and Chapter 4 continues with data preparation applying data mining methodology.

Chapter 4 Data preprocessing and selection

Raw sensor data that was collected from the sensors is described in Chapter 3. The importing and preparation of the sensor data into the software package Matlab is described in this Chapter.

This Chapter follows general guidelines on data mining methods (Kantardzic, 2011; Myatt & Johnson, 2009; Myatt, 2007) regarding the pre-processing of raw sensor data:

1. Import data
2. Characterise data and value mapping
3. Cleaning of data
4. Remove variables
5. Transform data

This Chapter presents the data import, the creation of abstract sensors, the initial data explorations, and concludes with the choice of sensors used in initial prediction models.

4.1 Data import

The recorded data was saved in different locations and formats. The data was collected on a computer which also ran the software package Matlab, as can be seen in fig. 3.7 on page 26.

The Matlab software package from MathWorks was selected because of its numerous methods for modelling techniques, ranging from linear models to non-linear neural network models. Matlab was also chosen, because of its fast vector and matrix processing, compared to traditional programming languages, for example C/C++. The fast vector and matrix calculations were necessary due to the large amount of data that was going to be worked on. The ability to access data from different sources, such as Structured Query Language (SQL) and Comma Separated Value (CSV) files, was a further consideration when choosing Matlab.

During the data import into Matlab, the name of each sensor was determined as follows:

- For SQL database sensors, an 'S' followed by the unique ID number and the location notes were used. For example, the sensor 15 had as location notes 'Ground air' and was therefore imported as *S15_GroundAir*.
- The web scraped sensors had the filename as a prefix, followed by an individual column heading that represented the collected data and unit. For example, the outside air temperature from the Chimet website was measured in °C and recorded in the file 'ChimetOutput.txt'. Therefore the imported sensor variable was named *ChimetOutput_temp_C*.

- Variables names were not permitted to contain spaces, therefore underscores were used to join name elements or camel case was used.

4.1.1 Discarded data

The sensor data recorded in the SQL database and CSV files was filtered during the importing process. The affected sensors are named in this Subsection and an exclusion reasoning is given.

The SQL database recorded data that was used to monitor the outside pool temperatures, but as they were not in close proximity to the house, those were excluded from pre-processing and modelling:

Table 4.1: *Discarded 1-wire network sensors*

Sensor names
S1_EnvironmentallySealedSensor
S2_GroundSensorUnderPool0x28belowInsulation0x29
S3_FloorSensorBelowPool0x28aboveInsulation0x29
S5_GroundCenterPool0x28noInsulation0x29
S6_FloorCenterPool0x28noInsulation0x29
S7_PoolWaterTemp
S8_CoverStatusSwitchGreyBox

The BBC website readings for three day forecasts were based on Met Office data and therefore redundant, except for daylight hours, pressure and humidity:

Table 4.2: *Discarded BBC sensor*

Sensor names
BBCOutput_max_temp_C
BBCOutput_min_temp_C
BBCOutput_uvrisk
BBCOutput_wind_direction
BBCOutput_wind_speed_miles

4.1.2 Error values from sensor readings

The 1-wire sensors collected erroneous data, which was filtered before further processing. The temperature value ‘85’ was returned when a temperature sensor had issues with the voltage supply. The temperature value ‘85’ was therefore deleted from all sensors with following conditions:

1. If the overall maximum value of the sensors was ‘85’, then delete all occurrences of ‘85’.
2. If the maximum was not equal to a value of ‘85’ and deletion of ‘85’ values only impacted 1% of the data, delete all occurrences of ‘85’. The percentage was selected to ensure that real temperature values, such as collected from sensors attached to the boiler pipes, were not affected.
3. Or else, Keep all data.

Sensors attached to the boiler pipes were affected by condition number two, deleting 0.07% of the recorded data. The first occurrences of '85' were found during the retrieval and inspection of the recorded data. The connections of the affected sensors and soldering methods of sensors were re-examined at that point and tested in lab conditions, before they were applied. Future occurrences of '85' were then used as an indicator for faulty sensor connections. Other possible errors were caught by the HA7net internally and not recorded as valid data.

4.1.3 Value mapping and adjustment of data

Sensor data was saved in a data structure in Matlab, one structure storing data from files and a second structure storing data from the SQL database. Each of the sensors was then saved as a matrix table as follows:

- A normal sensor, for example temperature sensor, had two columns: one for the timestamp value and one for the data values.
- A forecasted sensor, for example a forecasted temperature had three columns: one for the timestamp when the data was retrieved, one for the forecasted date value and one for the data value.

The 1-wire sensor numerical values were decompressed during the import process, as these were saved in a compressed state in the SQL database. The compression algorithm updated a timestamp of the last of two consecutive readings if they had the same value. If the time difference between those two readings was greater than one hour, a new value was saved. The time difference of at least hourly recorded data was selected to ensure that error checking was still possible. The diagram in fig. 4.1 shows the compressed data and the decompressed data after reconstruction, where the *RefreshTime* from the *Sensors* SQL table was used.

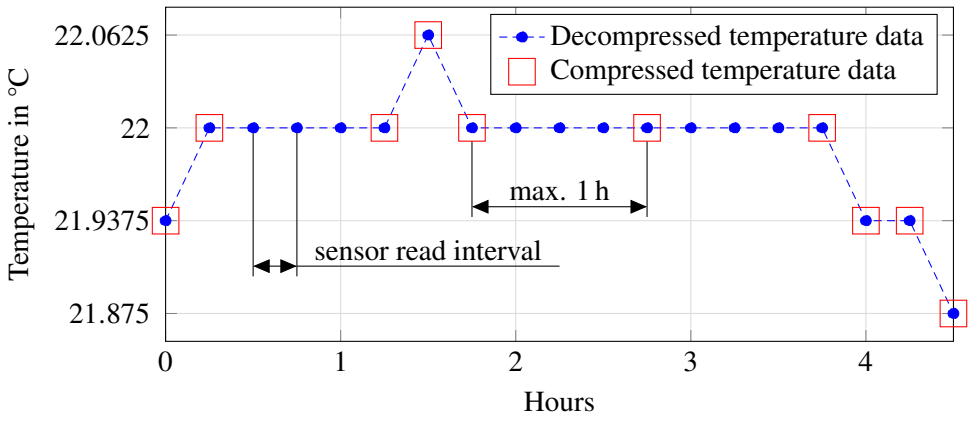


Figure 4.1: *Example data decompression showing decompressed and compressed data*

Data collected from the BBC and Met Office web pages included categorical or symbolic data for measurements of wind directions, wind speeds, UV risk and visibility. The following list shows the original and dummy conversions that were used to create sensors with numerical data values:

- Wind directions: each of the data values recorded were strings from the list: 'N', 'NNE', 'NE', 'ENE', 'E', 'ESE', 'SE', 'SSE', 'S', 'SSW', 'SW', 'WSW', 'W', 'WNW', 'NW', and

'NNW'. These categorical values were converted to numerical angular values in degrees, with a resolution of 22.5° (0° for 'N' to 337.5° for 'NNW').

- Wind speeds: wind speeds recorded 'Calm' converted to $0\,\mathrm{km\,h^{-1}}$.
- UV risk 'low', 'moderate' to unit-less values of '1' and '2'.
- Visibility was saved as strings 'Very Poor', 'Very Good', 'Poor', 'Moderate', 'Good', and 'Excellent' were converted to unit-less values of '1', '2', '4', '8', '16' and '32' respectively.

4.1.4 Aggregated sensors

The sensor data was used to construct new abstract sensors. The origin of each of these abstract sensors and their values are described in this Section.

4.1.4.1 BBC daylight hours

Daylight hours were derived from the sunrise and sunset data provided by the sensors *BBCOutput_sunrise_date* and *BBCOutput_sunset_date*. The amount of daylight hours was calculated as follows:

$$daylight\ hours = BBCOutput\_sunset\_date - BBCOutput\_sunrise\_date \tag{4.1}$$

These were then saved in *BBCdaylightHours* as a matrix with two columns: one for a timestamp and one for a daylight value in hours.

4.1.4.2 Remaining day/night hours

The data saved in *BBCdaylightHours* did not include the length of remaining day or night at a certain time of day. The values for sunrise and sunset times were taken from the BBC *sunset* time and the next day's *sunrise* time and *startDate* was the date and time at which the remaining day or night hours were to be calculated from. The variable *remainingDayNightHours* was then calculated with following rules:

$$remaining = \begin{cases} sunset - startDate & \text{if } sunset > startDate > sunrise \\ startDate - nextSunrise & \text{if } startDate > sunrise \text{ and } startDate > \text{sunset} \\ startDate - sunrise & \text{if } startDate < sunrise \end{cases} \tag{4.2}$$

Using this equation resulted in negative values for remaining night hours and positive values for remaining day light hours. The variable was saved as a two column matrix, a timestamp and the day/night hour value.

4.1.4.3 Vacant space

The alarm status sensor collected data from a 1-wire switch sensor which detected vacancy and occupancy states. The numerical values and their interpretation for the alarm status sensor were as follows:

4 normal state

6 exit horn on

12 alarm set

14 alarm set and exit horn on

A value of '12' and higher was interpreted as 'vacant' and lower than '12' as 'non-vacant'. The *vacant_space* abstract sensor was then created with following rules:

$$vacant\_space = \begin{cases} 1 & \text{if } S31\_AlarmStatusSensor \geq 12 \\ 0 & \text{if } S31\_AlarmStatusSensor < 12 \end{cases} \tag{4.3}$$

This abstract sensor used two columns: timestamps and a binary value for vacancy, with '1' for vacant and '0' for non-vacant.

4.1.4.4 Boiler on

Sensors installed on the pipes of the boiler recorded the water temperatures for the central heating flow, central heating return, cold water in and hot water out. A threshold temperature value was used to infer if the residential home was being heated or not. The abstract sensor *boilerOn* used the temperature of the sensor *S18_BoilerCHFlow*. A temperature threshold of 40 °C was selected. Temperatures for this sensor did not exceed 30 °C when turned off, therefore 40 °C was sufficient to determine times of heating for the boiler. The *boilerOn* sensor was derived as follows:

$$boilerOn = \begin{cases} 1 & \text{if } S18\_BoilerCHFlow \geq 40\,^{\circ}\text{C} \\ 0 & \text{if } S18\_BoilerCHFlow < 40\,^{\circ}\text{C} \end{cases} \tag{4.4}$$

The sensor was saved as timestamp and a binary value for boiler on, with '1' for boiler on and '0' for boiler off.

4.1.4.5 Heating demand

The heating demand was abstracted from *S32_HeatingZones* which interfaced the heating control system with a 1-wire sensor, described in Subsection 3.4.1 on page 26. A value greater than zero was interpreted as a heating demand. The new abstracted sensor *heatingDemand* was created with the following rules:

$$heatingDemand = \begin{cases} 1 & \text{if } S32\_HeatingZones > 0 \\ 0 & \text{if } S32\_HeatingZones = 0 \end{cases} \tag{4.5}$$

4.1.4.6 Heated house

A further approach determined the heating state and used two aggregated sensors: *boilerOn* and *heatingDemand*. Combining both sensors resulted in the creation of *HeatedHouse*.

$$HeatedHouse = \begin{cases} 1 & \text{if } heatingDemand = 1 \text{ OR } boilerOn = 1 \\ 0 & \text{if } heatingDemand = 0 \text{ AND } boilerOn = 0 \end{cases} \tag{4.6}$$

This aggregated sensor ensured, that times of the heating 'off' state were recorded only if the heating demand was 'off' and the central heating water temperature was below the threshold of 40 °C.

4.1.4.7 Vacant cold and warm

The abstract sensor *vacant_space* indicated vacancy but did not indicate a heating state, as both a heated and unheated vacancies were possible. The sensors *vacantColdSpace* and *vacantWarmSpace* split up the heating states with *HeatedHouse* and *vacant_space* as follows:

$$vacantColdSpace = \begin{cases} 1 & \text{if } HeatedHouse = 0 \text{ AND } vacant\_space = 1 \\ 0 & \text{if } HeatedHouse = 1 \text{ OR } vacant\_space = 0 \end{cases} \tag{4.7}$$

$$vacantWarmSpace = \begin{cases} 1 & \text{if } HeatedHouse = 1 \text{ AND } vacant\_space = 1 \\ 0 & \text{if } HeatedHouse = 0 \text{ OR } vacant\_space = 0 \end{cases} \tag{4.8}$$

4.1.4.8 Night time

The vacancy sensor did not yield a large number of long vacancy periods, therefore night time was considered. The sensor *boilerOn* was used to distinguish between heated and unheated states during night time. Furthermore, night times were also considered, because of decreased dynamics from inhabitants or changes in doors and windows open/closed state, and no solar influences. The night time data was used to extend datasets which had no vacancy detection due to the lack of necessary sensors or to create datasets similar to vacancy. Night time was defined as the time between 11 o'clock at night and 8 o'clock in the morning and saved temporarily as variable *nightTime*, which was then used to derive two further abstract sensors, inhabited nights with heating and without heating.

$$inahbitedNightWarm = \begin{cases} 1 & \text{if } HeatedHouse = 1 \text{ AND } nightTime = 1 \\ 0 & \text{if } HeatedHouse = 0 \text{ OR } nightTime = 0 \end{cases} \tag{4.9}$$

$$inahbitedNightCold = \begin{cases} 1 & \text{if } HeatedHouse = 0 \text{ AND } nightTime = 1 \\ 0 & \text{if } HeatedHouse = 1 \text{ OR } nightTime = 0 \end{cases} \tag{4.10}$$

4.1.4.9 Hourly forecasts

The Met Office forecasts provided hourly forecasts for up to 24 hours into the future. The forecasts were split up into forecast sensors for one, two, three, four, five, et cetera, up to 24 hours. This step was carried out to create forecast sensors with different prediction horizons. The forecasted hours were calculated by subtracting the access time of the data from the forecasted time. For example, if forecasted data was collected at 2pm and the forecasted time was 3pm, then the data would be saved under a forecast for one hour. This step was carried out for all the data recorded in the Met Office forecast sensors and then the data was saved in 24 separate sensors for each weather

information, such as temperature and wind speed.

$$\triangle forecastedHours = forecastTime - accessTime \tag{4.11}$$

4.1.4.10 Daily forecasts

The Met Office forecast provided daily predictions, and these were split up into one, two, three and four day forecast sensors. As each forecast contained a timestamp and forecast date, the time difference was used to put forecasts into the correct category. The calculation was similar to the hourly forecasts, replacing hours with days.

$$\triangle forecastedDays = forecastDay - accessDay \tag{4.12}$$

4.1.4.11 Forecast trending

The trending of forecasts was used to generate abstract sensors. Gradients of temperature forecasts were generated with a linear regression calculation, which returned results in a $°C\,h^{-1}$ (degree Celsius per hour) and a $°C\,d^{-1}$ (degree Celsius per day) gradient that described temperature trends. For example, a value of $-1\,°\,h^{-1}$ was interpreted as a decrease of 1° per hour. The general formula of the linear fit was:

$$y(x) = ax + b \tag{4.13}$$

The trending of forecasts was extracted from Met Office temperature forecasts, as they provided both hourly and daily predictions. A first order linear regression fit was applied to a set of temperature values recorded at the same time. The gradient a of the linear fit was then saved as the trending value for daily forecasts from one to four days in degrees per day and for hourly forecasts from one to 12 hours in degrees per hour. The abstract sensors were *metOutput_temp_C_trend_12hour* and *metOutput_temp_C_trend_4day*. An example of a 12 hour trend can be seen in fig. 4.2. The plot shows the original data for each forecasted hour from one up to 12 hours and a linear fit. The highlighted value in the legend on the top right was the gradient a, which was saved in *metOutput_temp_C_trend_12hour*.

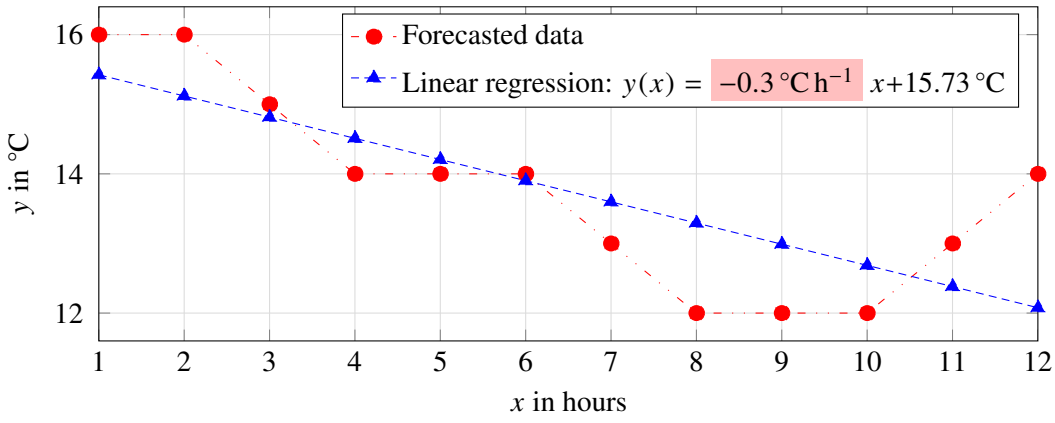


Figure 4.2: *Example of a linear fit with a temperature gradient created from observed temperature data*

A numerical solution (Weisstein, 2013a) for the trending parameter a was defined as follows:

$$a = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{j=1}^{n} y_j}{\sum_{i=1}^{n} (x_i^2) - \frac{1}{n} (\sum_{i=1}^{n} x_i)^2} \tag{4.14}$$

The parameters x_i and $y_{i,j}$ were the values for time and temperature, and n was the number of values. In the example plot, n was 12 for *metOutput_temp_C_trend_12hour*.

4.1.4.12 Improved heating state

The accuracy of the heating state was essential in the improved model, see Chapter 7. The previously used *HeatedHouse* sensor was reviewed since heating on/off states did not match with responses recorded from air temperature sensors in the residential home. A visual analysis of plots showing both heating state and landing air temperature revealed that the heating off state was recorded later than the actual influence observed for the landing air. The landing air temperature decline happened earlier, with time differences from five minutes up to one hour compared to the actual heating off state detected by *HeatedHouse*. Therefore the heating state aggregated sensor algorithm was redefined. An improved version was created, which took into account that the observed boiler central heating water temperature had a faster decline when turned off and a quicker rise when turned on. The new approach therefore used two different temperature thresholds instead of one. The temperatures selected were 30 °C for a heating 'on' and 55 °C for a heating 'off' state detection rule. An example plot, showing the landing air temperature and the old and new aggregated *HeatedHouse* can be seen in fig. 4.3. The plot shows the landing air temperature for one day and the states for the old and new aggregated *HeatedHouse*. The plot shows, that the old *HeatedHouse* did not immediately indicate the heating 'off' state. The new *HeatedHouse* aggregated sensor included a higher number of shorter heating on/off cycles, compared to the old aggregated sensor. The new aggregated sensor replaced the old *HeatedHouse* for the improved model creation and data selection process in Chapter 7.

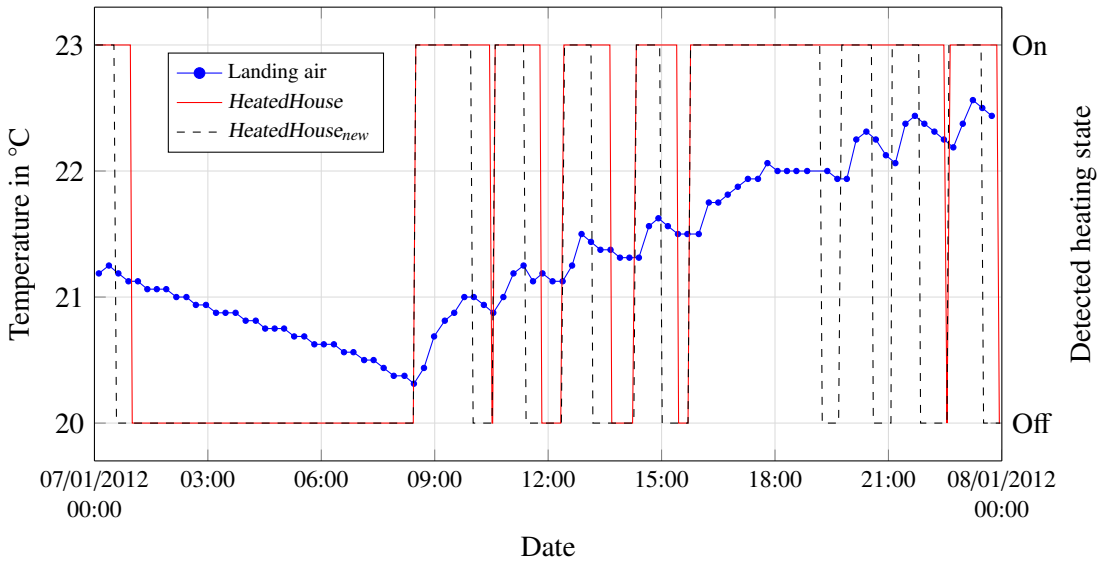


Figure 4.3: *Landing air temperature sample with scaled old and new aggregated sensor HeatedHouse*

$$HeatedHouse_{new} = \begin{cases} 1 & \text{if } heatingDemand = 1 \text{ OR if } S18\_BoilerCHFlow \geq 30\,^{\circ}\text{C} \\ 0 & \text{if } heatingDemand = 0 \text{ AND if } S18\_BoilerCHFlow < 55\,^{\circ}\text{C} \end{cases} \quad (4.15)$$

4.1.4.13 Discussion

The collected sensor data from the residential home was used to analyse and create new virtual sensors that accessed 'hidden' information within the sensor data. The sensor data was used to create new sources of information by combining sensors to create new aggregated sensors. Examples of aggregated sensors can be seen in heating and vacancy state sensors.

4.1.5 Final imported data structure

The structure *sensors* held the data for the imported and aggregated sensor data. The sensor names can be seen in table 4.3.

Table 4.3: *Names of imported and aggregated sensors*

Sensor names	
S9_EastWallOuterSensor	*vacantCold*
S10_EastWallInnerSensor	*vacantWarm*
S11_FirstFloorCoreTemp	*inhabitedNightCold*
S12_SilverTubeAirTemp	*inhabitedNightWarm*
S13_FirstFloorLandingAir	*vacantColdNightCold*
S14_GroundEastInnerWall	*remainingDayNightHours*
S15_GroundAir	*BBCOutput_pressure_mbar_1day*
S16_BoilerColdWater	*BBCOutput_pressure_mbar_2day*
S17_BoilerHotWater	*BBCOutput_humidity_rel_percent_1day*
S18_BoilerCHFlow	*BBCOutput_humidity_rel_percent_2day*
S19_BoilerCHReturn	*metOutput_temp_C_1day*
S20_EavesN0x2DEAir	...
S21_RoofUnderTilesN0x2DEEaves	*metOutput_temp_C_4day*
S22_RoofUnderTilesN0x2DMidEaves	*metOutput_temp_C_trend_4day*
S23_EavesN0x2DMidAir	*metOutput_wind_direction_1day*
S24_EavesWallN0x2DMidRoomSideRad	...
S25_EavesWallN0x2DMidEavesSide	*metOutput_wind_direction_4day*
S26_ComfortUser1Hall	*metOutput_wind_speed_miles_1day*
S27_ComfortUser2Hall	...
S28_ComfortUser3Hall	*metOutput_wind_speed_miles_4day*
S29_ComfortUser4Hall	*metOutput_visibility_1day*
S30_ComfortUser5Hall	...
S31_AlarmStatusSensor	*metOutput_visibility_4day*
S32_HeatingZones	*metOutput_temp_C_1hour*
S34_RoofAir	*metOutput_temp_C_2hour*
S35_RoofApex	...
S36_GroundEastInnerWall2	*metOutput_temp_C_24hour*
S37_GroundAir2	*metOutput_temp_C_trend_12hour*
smart_temp_C	*metOutput_wind_direction_1hour*
smart_humidity_rel_percent	*metOutput_wind_direction_2hour*
smart_wind_speed_km	...
smart_wind_direction	*metOutput_wind_direction_24hour*
smart_solar_radiation_W_per_m2	*metOutput_wind_speed_miles_1hour*
smart_rain_rate_mm	*metOutput_wind_speed_miles_2hour*
smart_pressure_mbar	...
ChimetOutput_temp_C	*metOutput_wind_speed_miles_24hour*
ChimetOutput_pressure_mbar	*metOutput_visibility_1hour*
BBCdaylightHours	*metOutput_visibility_2hour*
vacant_space	...
heatingDemand	*metOutput_visibility_24hour*
boilerOn	*HeatedHouse*

4.2 Cleansing data

The raw sensor data contained erroneous data records and random outliers that might have had an impact on modelling. Although value errors of '85' were dealt with, other errors were still possible, for example temperature value outliers. Common preventive measures are listed below, which ranged from data exclusion to different detection methods and dealing with outliers. The measures taken to address the issue of erroneous data were to delete, replace or adjust the affected data.

4.2.1 Excluding data

After all sensor data was imported and the aggregated sensors created, data was segmented and data was excluded. Reasons for these methods were as follows:

- The data from an experiment was not included as it biased the recorded data. The experiment was carried out between the 21st and 27th October 2011 in the residential home and is described in Chapter 6. The recorded experimental data was different from day to day data because of prolonged continuous heating and continuous cooling.
- The residential home underwent structural changes which affected heat distribution and dynamics in day to day data, if temperature sensors were at the same locations. The structural changes were introduced after the 27th of October. Therefore different subsets were created. A new boiler was introduced on the 21st October 2012, marking the end of a dataset starting from the experiment.

4.2.2 Winsorising

The winsorising method limited outlier data with statistical minimum and maximum values. If the data was a Gaussian normal distribution, Winsorising with a setting of 90% would replace values below 5% with the minimum value and replace values above 95% with the maximum value. For example: data was recorded in the range of -5 to +5 and 90% of the data was within a range of -3 to +3. Values smaller than -3 were replaced with -3 and values above +3 were replaced with +3.

The recorded data from the residential home included information, which could include extreme values. The method of winsorising provided a replacement value for outliers, which could impact extreme values. Winsorising was therefore not selected to replace outliers within the work described in this Dissertation.

4.2.3 Sigma value trimming

In a Gaussian normal distributed population, certain values lay within a certain percentage at the higher and lower end of the population mean. The standard deviation σ was used as a measure to define percentage limits. Table 4.4 shows the different percentages of sigma values and the respective data within this range for normally distributed data. Figure 4.4 shows a plot of a standard normal distribution derived from Weisstein (2013b). Values of 6σ and higher were sufficient to withhold the important data and exclude outliers, that would not fit the data.

Table 4.4: *Sigma values and corresponding percentages of data within the limits of the sigma values for normal distributed data*

$z\sigma$	Percentage[1]
1σ	68.269 949
2σ	95.449 974
3σ	99.730 020
4σ	99.993 666
5σ	99.999 943
6σ	99.999 999

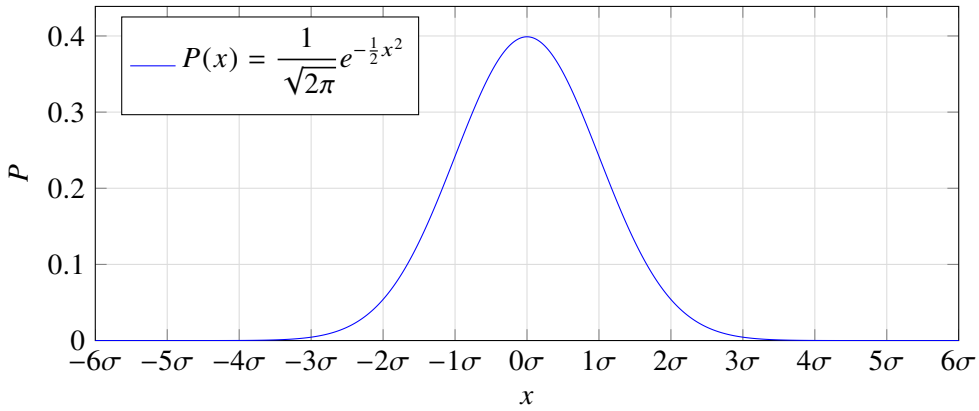


Figure 4.4: *Standard normal distribution curve*

4.2.3.1 Discussion

The sigma trimming method deleted outliers dependant on the populations mean and standard deviation. Sigma trimming could not be applied to categorical data, such as wind directions, thermal comfort input and binary sensors, because it would delete whole categories of values. Therefore it was decided to exclude the categorical sensors from trimming.

In this Section, outliers were addressed with 6σ. Visual confirmations for sensors affected by sigma trimming were plotted to ensure that only outliers were deleted and not relevant data.

Sensors affected by sigma value trimming can be seen in table 4.5.

Table 4.5: *Sensors and affected sets by data cleansing*

Sensor name	Affected values
S10_EastWallInnerSensor	7
S11_FirstFloorCoreTemp	3
S17_BoilerHotWater	3
S24_EavesWallN0x2DMidRoomSideRad	16

4.3 Transforming data

4.3.1 Normalization of data

The imported sensor data and aggregated sensors were translated to or collected in numerical values which had different scales and did not have a common value range. The application of a scaling method for all collected sensor values to a common range enabled the comparison of the different sensor data. Three normalization methods are presented in this Section.

4.3.1.1 Min-Max Normalizing

A simple normalizing function was min-max normalization. New minimum and maximum values were chosen to convert values to a new range:

$$x_{new} = \frac{x - min(x)}{max(x) - min(x)} * (max_{new} - min_{new}) + min_{new} \tag{4.16}$$

For example, a value of 50 within the range [40,120] was transformed to the range [0,1]:

$$x_{new} = \frac{50 - 40}{120 - 40} * (1 - 0) + 0 = 0.125 \tag{4.17}$$

4.3.1.2 Standardizing - z-score

Normalizing used the standard deviation as a benchmark and scaled the data to a standard deviation, this was also known as 'z-score':

$$x_{new} = \frac{x - \mu}{\sigma} \tag{4.18}$$

The μ was the mean value and σ was the standard deviation of the data. The resulting x_{new} was unit-less. For example, a value of 50 within the range [0,70], a mean value of 30 and a standard deviation of 5:

$$x_{new} = \frac{50 - 30}{5} = 4 \tag{4.19}$$

4.3.1.3 Decimal scaling

Transformations that used a decimal scale changed data to a range of -1 to 1. Decimal scaling used the absolute maximum of the selected data to determine the scaling factor as number of decimal places n.

$$x_{new} = \frac{x}{10^n} \tag{4.20}$$

For example, a value of 50 within the range [0,900] would result in n to be selected as 3 to describe the whole range.

$$x_{new} = \frac{50}{10^3} = 0.05 \tag{4.21}$$

4.3.2 Application of normalisation to sensor data collected

Normalisation criteria applied to data were saved for future normalisations and denormalisations. For example, when a min-max normalization was used, saving the max and min values and applying them to new values ensured the same range of normalised values. An alternative important use case for preserving the scaling method and parameters was to transform predicted data back for plots and visual confirmation of data.

The min-max and decimal scaling transformed the data onto a different scale and z-scores made the data unrecognisable to human interpretation. Furthermore the min-max used minimum and maximum values, which could be outliers and would skew the data. Applying a pre-processing step that would smooth out or delete extreme values would circumvent this. The min-max normalization was therefore applied to the sensor data with a range between 0 and 1 after outliers were treated with the sigma trimming method.

4.4 Segmentation of data

The collected data was segmented by dates, which were brought about by structural changes or sensor availability to record crucial information, such as the alarm sensor. To make sensors comply with the condition 'unheated vacant space', start and stop times were defined with the following constraints:

- Vacant space, that was when the alarm system was 'on'.
- A cold space, which was derived by looking at the boiler water temperature.
- Minimum periods of at least one hour were selected to ensure that multiple points of data were recorded.

The aggregated sensor *vacantColdSpace*, described in Subsubsection 4.1.4.7 on page 41, was used to produce start and stop times when the sensor had a value of 1, indicating vacancy and cold space. The minimum length for these cold vacancies was set as one hour. Applying this method yielded 161 distinct start and stop time sets from the 1st of April to the 21st of October 2011. A frequency diagram of dataset hours and can be seen in fig. 4.5 and a distribution of the dataset length over dates for the first set in fig. 4.6. Further analysis revealed, that the vacancies were not evenly spread throughout the week, see fig. 4.7. The distribution of vacancies was attributed to the inhabitants' use of the property.

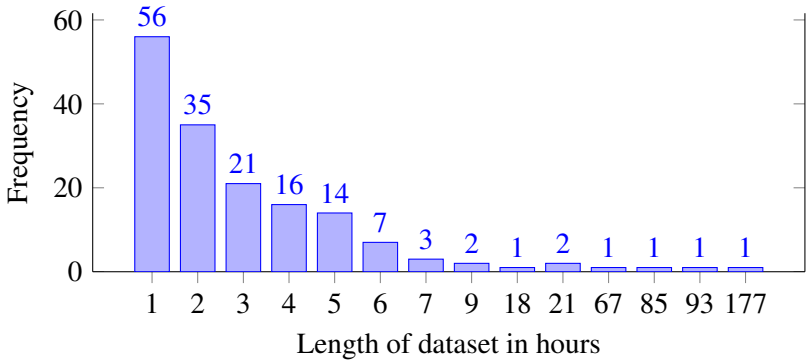


Figure 4.5: *Frequency distribution of dataset lengths in hours*

The data set was called the 'Golden dataset', which removed some dynamic inputs to the residential home. For example, inhabitants' movements, closing/opening of windows or the use of heat generating appliances.

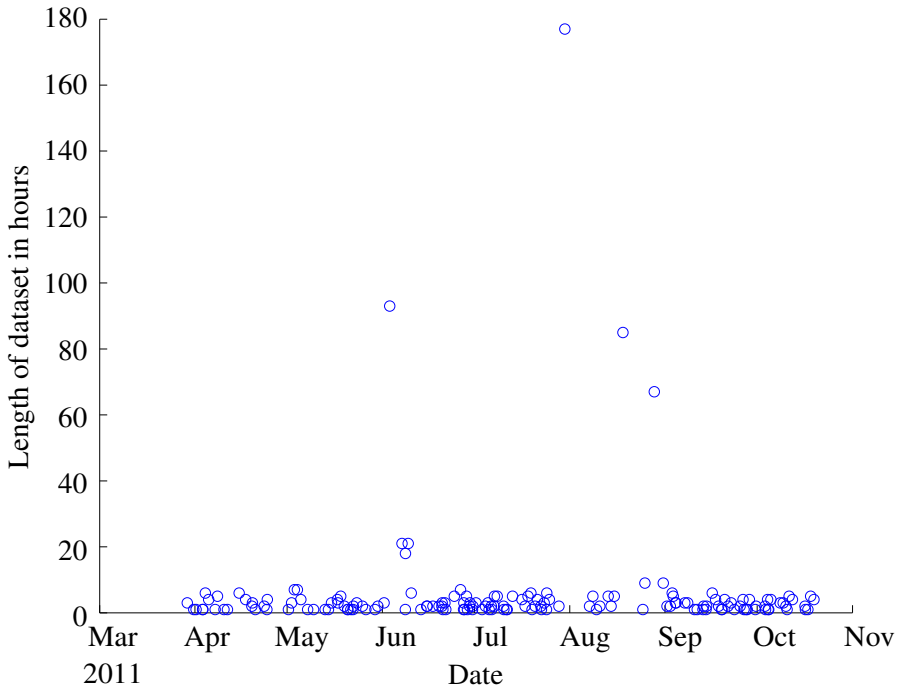


Figure 4.6: *Dataset length distribution over time*

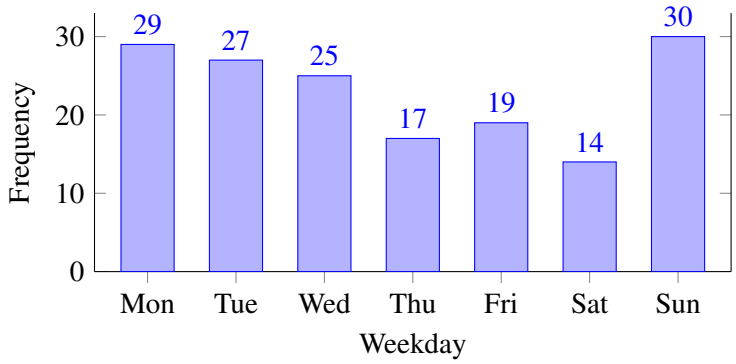


Figure 4.7: *Frequency distribution of datasets per weekday*

4.4.1 Building datasets

Based on the start and stop times from the aggregated sensor *vacantColdSpace*, observation data from all sensors was collected from the 'Golden dataset'. Initially, data recorded at the start time $t_0 = 0$ of a prediction was saved as 'start conditions'. The 'start conditions' were used as prediction model inputs. With the definition of an hour as the prediction minimum interval, 'start conditions' were saved for every beginning of an hour in multiple hour spanning datasets defined by *vacantColdSpace*. The reasoning was that if there was a dataset with x hours, multiple sets for prediction were generated from it. This approach expanded the chosen datasets for training and validation. The expansion of datasets can be seen in an example shown in fig. 4.8. The original four hour long vacancy is shown on the top and expansions for each shorter hourly prediction are shown below.

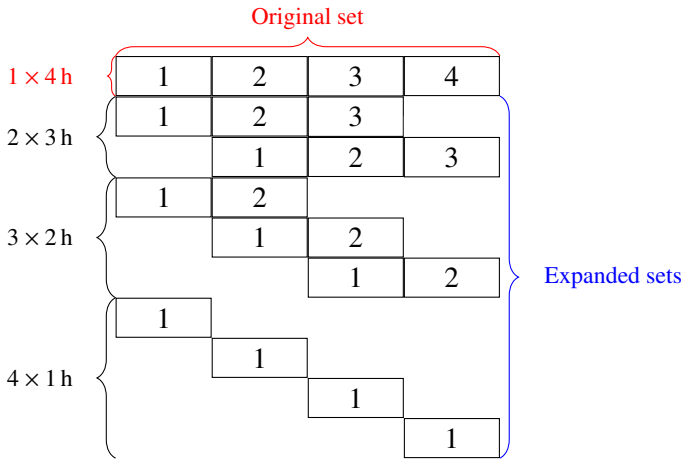


Figure 4.8: *Step by step expansion of an exemplary four hour 'original set'*

The matrices used as inputs and outputs for the models were created to keep data compressed, that was, the expansion for each smaller hour was carried out in place. For example: a four hour long output had also a three, two, and one hour set, sharing the same starting conditions from the inputs. With this approach, the number of expansions shown in fig. 4.8 was achieved, but a shorter matrix design was retained. Following are the matrix designs used for inputs and outputs respectively:

$$\text{inputs}=\begin{pmatrix} i_{11} & i_{21} & i_{31} & \dots & i_{p1} \\ i_{12} & i_{22} & i_{32} & \dots & i_{p2} \\ \dots & & & & \\ i_{1m} & i_{2m} & i_{3m} & \dots & i_{pm} \end{pmatrix} \quad \text{outputs}=\begin{pmatrix} o_{11} & o_{21} & o_{31} & \dots & o_{n1} \\ o_{21} & o_{31} & o_{41} & \dots & 0 \\ \dots & & & & \\ o_{n1} & 0 & 0 & \dots & 0 \\ o_{12} & o_{22} & o_{32} & \dots & o_{n2} \\ o_{22} & o_{32} & o_{42} & \dots & 0 \\ \dots & & & & \\ o_{n2} & 0 & 0 & \dots & 0 \\ \dots & & & & \\ o_{nm} & 0 & 0 & \dots & 0 \end{pmatrix}$$

p was the number of inputs, n the number of hours for a particular set and m the number of total expanded datasets. A detailed explanation of the matrices as follows:

- The start and stop times of a unheated vacancy were collected from the sensor *vacantColdSpace*.
- A first vacancy set was selected from the 'Golden dataset'.
- The first expanded set of inputs was collected at time $t = 0$, namely $i_{11}, i_{21}, \dots, i_{p1}$. The corresponding outputs for the different hourly models were $o_{11}, o_{21}, \dots, o_{n1}$.
- The second expanded set of inputs was collected at time $t = +1h$, namely $i_{12}, i_{22}, \dots, i_{p2}$. The corresponding outputs for the different hourly models were $o_{21}, \dots, o_{(n-1)1}$. The outputs

were effectively copied from the first set, shifted by one to the left and reducing the number of outputs by one for the longest hour.

- This expansion continued until the last expanded set of inputs was $i_{1n}, i_{2n}, \ldots, i_{pn}$ and the only output was o_{n1}, which was used for a one hour prediction.
- The second vacancy set was selected and the data expansion applied.
- This was continued until the last vacancy set.

4.5 Initial data explorations

The data pre-processing step resulted in the creation of data matrices for input variables and the output variable. For the output variable, the first floor landing air temperature was selected as an indicator for the indoor general air temperature. The ground floor was not selected due to sensors being exposed to heat radiation from a radiator, solar radiation influence from windows and through the front door, and drafts through the letter box. The second floor was not selected due to having only bedrooms, therefore not living space that would be used immediately after inhabitants return to the premises.

4.5.1 Datasets

The datasets created for the Golden dataset yielded 161 datasets, which included long vacancies of up to 170 hours. The few long vacancies would have skewed modelling processes, as the majority of vacancies observed were shorter. Therefore these long sets were regarded as holidays and discarded.

The landing air temperature sensor recorded a maximum increase of 0.5 °C and a maximum decrease of 0.5 °C during vacancies. For example if at the start of a vacancy the landing air temperature was 19.5 °C, the end temperature would be 20 °C if the temperature increased, and 19 °C if the temperature decreased. The resolution of the temperature sensor was 0.0625 °C therefore the maximum temperature difference was within eight digital steps of the temperature sensor. The core temperature showed even smaller temperature differences of 0.25 °C to −0.375 °C during periods of vacancy.

4.5.2 Correlation and redundancy of data

Correlation analysis was used to evaluate and compare sensor data. It incorporated visual and calculation based measures to estimate the correlation of two variables. The first method used scatter plots, where two variables were plotted. The resulting graph was able to show if a relationship was linear. The second method was to calculate a correlation coefficient r as a linear relationship measure (Myatt, 2007):

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \tag{4.22}$$

The two variables in this formula are x and y. The individual values for x are x_i and the individual values for y are y_i. $\bar{x}$ and $\bar{y}$ are the mean values of the respective variables. The number of observations is n, and s_x and s_y are the standard deviations for the respective variables. The

coefficient of correlation could identify a linear relationship between two variables with a number ranging between -1 and +1. A correlation coefficient around -1 was interpreted as a negative linear relationship, values near +1 as a positive linear relationship and values around 0 were identified as having no linear relationship. The r value did not indicate if any non-linear relationships existed or not.

The collected data from disparate sources and sensors in different physical locations collected similar data, for example air temperature, atmospheric pressure, wind speed and wind direction. The Chimet buoy was close to the residential home and a 3rd party source of information, which was selected to replace the weather station placed on the rooftop. A correlation analysis was used to confirm, if the Chimet buoy data was similar to the weather station data.

A scatter plot was used to plot measurements of the pressure and air temperature measurements from the on-site weather station and the Chimet buoy. Each data point in the scatter plot was recorded at the same time. Additionally the correlation coefficient r was calculated and is shown on both plots in figs. 4.9 and 4.10.

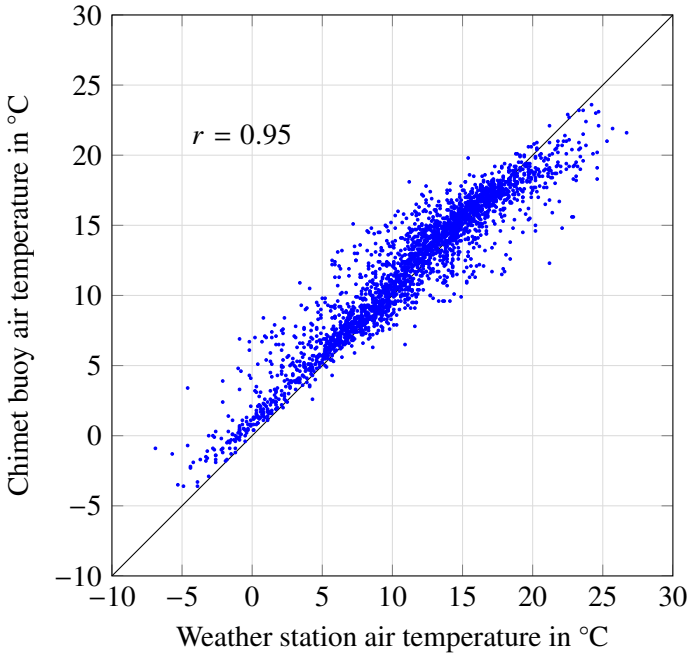


Figure 4.9: *Weather station and Chimet temperature measurement correlation*

The correlation coefficients were close to +1, indicating a positive linear correlation between the chosen variables. The scatter plot of the pressure values showed identical values and the temperature scatter plot showed, that the relationship was not purely linear. For cold temperatures recorded at the residential home, slightly higher temperatures were recorded at the buoy. This phenomenon was reversed in the summer, where colder temperatures were recorded at the buoy and warmer temperatures at the weather station. This effect could be explained with the buoy being located in the sea. The sea water temperatures changed slowly and therefore would increase the cold air temperature and decrease the hot air temperature. The residential home being 10 km inland did not experience these influences. Although the correlation for the temperature values was not purely

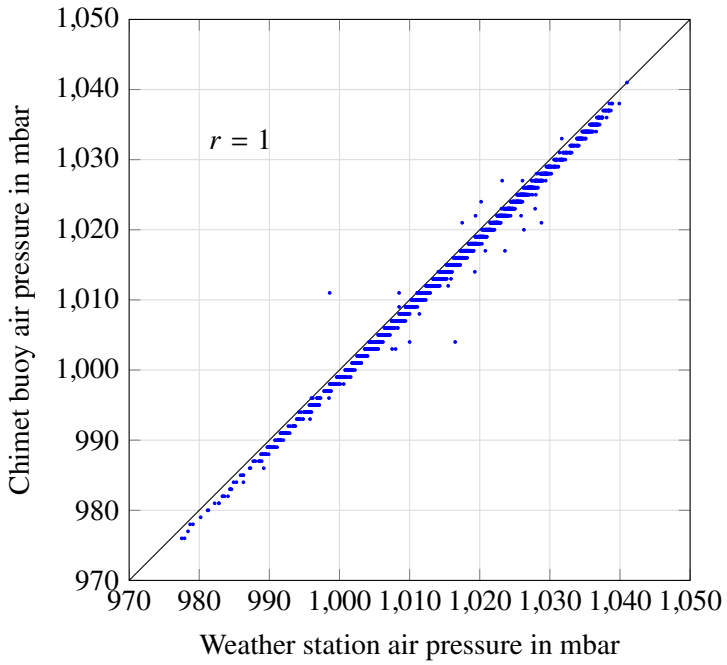


Figure 4.10: *Weather station and Chimet pressure measurement correlation*

linear, the investigated sensors from the on-site weather station were replaced with the sensors from the Chimet buoy. This decision was made to ensure, that models created from sensor data would include remote instead of local current weather information.

4.5.3 Evaluating and comparing website forecasts

The collected forecasts were split into hourly forecasts to use in later models. Both forecasts and local meteorological data was collected from websites, which were compared. In this Dissertation, temperature forecasts from the Met Office were compared with actual data from the Chimet website and on-site weather station. A maximum of 24 hour forecasts with one hour intervals was available. Table 4.6 shows the results for each hour and for the two comparison variables.

The weather forecast fitted the on-site weather station measurements closer than the readings from the buoy website. This was verified by comparing values from Mean Absolute Error (MAE), Mean Squared Error (MSE), Relative Absolute Error (RAE) and Relative Squared Error (RSE) results. The benchmark values were smaller for the weather station, than for the buoy. R^2 values indicated, that the on-site weather station was a better fit for the Met Office forecast than the Chimet buoy. Generally the values worsened for longer hourly forecasts, and this was seen as normal behaviour, because a longer period of time was attributed with increased variations in between that might influence a forecast. The Mean Absolute Scaled Error (MASE) values were worse for short term predictions in the range of a few hours. The results showed, that for short term predictions, a naïve forecast was more suitable than the Met Office forecast. Conclusions drawn from these comparisons were:

- Met Office forecast fitted better to the local weather station data than the Chimet buoy data.
- Short period forecasts were worse than naïve forecasts.

Table 4.6: *Efficiency measures for Met Office temperature hourly forecast versus observed data*

	Weather station						Chimet buoy					
h	R^2	MAE/°C	MSE/°C^2	RAE	RSE	MASE	R^2	MAE/°C	MSE/°C^2	RAE	RSE	MASE
1	0.95	0.98	1.81	0.19	0.05	1.50	0.93	0.99	1.88	0.22	0.07	2.50
2	0.94	1.06	2.15	0.21	0.06	0.95	0.94	0.96	1.74	0.22	0.06	1.52
3	0.93	1.13	2.43	0.22	0.07	0.73	0.94	0.95	1.68	0.22	0.06	1.14
4	0.93	1.15	2.52	0.23	0.07	0.60	0.94	0.95	1.68	0.22	0.06	0.96
5	0.93	1.15	2.58	0.23	0.07	0.50	0.94	0.99	1.79	0.22	0.06	0.85
6	0.93	1.12	2.46	0.22	0.07	0.42	0.93	1.02	1.87	0.23	0.07	0.77
7	0.93	1.13	2.47	0.22	0.07	0.39	0.93	1.02	1.87	0.23	0.07	0.73
8	0.94	1.11	2.41	0.22	0.07	0.35	0.93	1.04	1.93	0.23	0.07	0.68
9	0.94	1.11	2.39	0.22	0.06	0.33	0.93	1.06	1.98	0.24	0.07	0.65
10	0.94	1.10	2.37	0.22	0.06	0.32	0.93	1.05	1.98	0.24	0.07	0.64
11	0.94	1.10	2.36	0.22	0.06	0.31	0.93	1.05	1.98	0.24	0.07	0.62
12	0.94	1.11	2.37	0.22	0.06	0.31	0.93	1.07	2.03	0.24	0.07	0.60
13	0.94	1.12	2.40	0.22	0.07	0.31	0.93	1.07	2.01	0.24	0.07	0.60
14	0.94	1.12	2.41	0.22	0.07	0.31	0.93	1.07	2.01	0.24	0.07	0.61
15	0.94	1.12	2.41	0.22	0.06	0.32	0.93	1.08	2.04	0.24	0.07	0.61
16	0.94	1.12	2.41	0.22	0.07	0.34	0.93	1.08	2.02	0.24	0.07	0.62
17	0.93	1.13	2.45	0.22	0.07	0.35	0.93	1.08	2.04	0.24	0.07	0.64
18	0.93	1.14	2.49	0.22	0.07	0.38	0.92	1.08	2.08	0.25	0.08	0.66
19	0.93	1.15	2.52	0.23	0.07	0.41	0.92	1.12	2.34	0.25	0.08	0.71
20	0.92	1.20	2.78	0.24	0.08	0.47	0.90	1.18	2.66	0.27	0.10	0.77
21	0.92	1.27	3.10	0.25	0.08	0.54	0.87	1.32	3.43	0.30	0.13	0.89
22	0.92	1.28	3.13	0.25	0.08	0.59	0.88	1.30	3.23	0.29	0.12	0.90
23	0.92	1.28	3.15	0.25	0.08	0.62	0.88	1.32	3.38	0.30	0.12	0.92
24	0.91	1.30	3.27	0.25	0.09	0.66	0.87	1.34	3.49	0.30	0.13	0.92

- The mid-range Met Office forecasts performed better than naïve forecasts.

Daily forecasts up to 4 days were accessible from the Met Office website and these forecasts were compared to the on-site weather station and the Chimet buoy data. Results of this analysis can be seen in table 4.7.

Table 4.7: *Efficiency measures for Met Office temperature daily forecast versus observed data*

	Weather station						Chimet buoy					
d	R^2	MAE/°C	MSE/°C^2	RAE	RSE	MASE	R^2	MAE/°C	MSE/°C^2	RAE	RSE	MASE
1	0.80	2.08	7.75	0.41	0.21	1.03	0.89	1.32	3.01	0.30	0.11	0.92
2	0.74	2.39	9.85	0.47	0.26	0.92	0.84	1.60	4.32	0.36	0.16	0.85
3	0.75	2.38	9.45	0.46	0.25	0.82	0.84	1.64	4.39	0.37	0.16	0.78
4	0.55	3.48	18.63	0.66	0.48	1.12	0.59	2.95	13.55	0.65	0.48	1.30

The results were comparable with the hourly forecasts: the shorter the prediction horizon, the better the result. The MASE values had a local minimum at three days and, compared to results from hourly comparisons, showed a poorer performance in general. The lowest value for MASE was the three day forecast and the highest the four day forecast, however none reached the scores observed from the hourly forecasts. Compared with hourly forecasts, daily forecasts had a better fit for the Chimet buoy when considering all performance values.

The prediction performance values demonstrated, that forecasts were able to predict local weather at and close-by the residential home. Hourly forecasts of around 12 hours showed a higher accuracy than long term hourly forecasts. Daily forecasts increased accuracy compared to a naïve forecast, but were not as accurate as hourly forecasts.

The results concurred with the performance assessments from the Public Weather Service Customer Group (PWSCG). The performance assessments set probability target limits for different prediction lengths for the Met Office forecasts. The performance results showed (The Met Office, 2013b), that short term predictions for temperatures were up to 90% correct and up to two days within 75%, within an accuracy of ±2 °C.

4.5.4 Further segmentation of the data

The 'Golden dataset' covered a period of predominantly warmer weather conditions, therefore no winter conditions were observed. Further data recorded after the experiment was treated differently as the residential home was remodelled to an open plan layout on the ground floor level, which would impact indoor heat distribution.

Due to the limitations of the 'Golden dataset', the data was extended to include winter conditions. This was achieved by adding data from November 2010 until April 2011. As no indicator for vacancy or heating demand was available during this period, the aggregated sensor *inhabitedNightWarm* was used, described in Subsubsection 4.1.4.8 on page 41. Night time was used to ensure that the recorded data was not affected by inhabitants during dormancy. This simulated vacancy was then added to the 'Golden dataset' and the resulting dataset was called 'Golden Extended dataset'.

In October 2012, a new boiler was fitted in the residential home, leading to an additional cut-off date for the datasets. This created a third dataset range from November 2011 - the end of the experiment - until early October 2012 and was called 'After Experiment dataset'.

The segmented datasets are shown in table 4.8 with dates and the number of unheated vacancy sets. A plot depicting the dates can be seen in fig. 4.11.

Table 4.8: *Segmented sets created from historical data*

Name	Start date	End date	Days	Unheated vacancies
Golden	2011-03-28	2011-10-21	204	157
Golden Extended	2010-11-24	2011-10-21	331	455
After experiment	2011-11-01	2012-10-12	337	266

A detailed breakdown of the actual data per hour and the expanded values by dividing longer hourly sets into smaller hourly sets can be seen in fig. 4.12 on the following page and fig. 4.13 on the next page. The original datasets for x-hours are given in red above the bars and in blue representing the sum of expanded sets. For example in fig. 4.12 on the following page, two sets for 21 hours were available, with the expansion, there were four sets each 20 hours long, until there were 464 total sets of one hour long sets expanded.

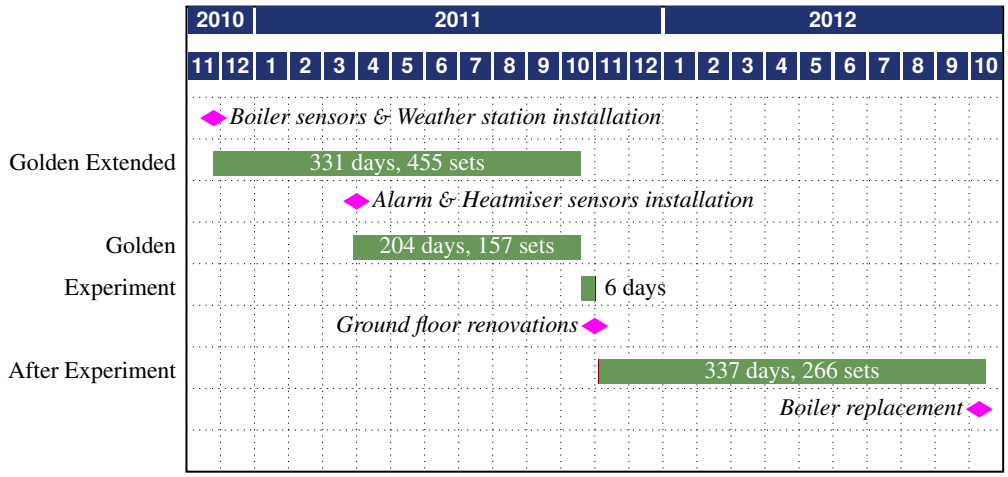


Figure 4.11: *Datasets and important events during the data collection in the residential home*

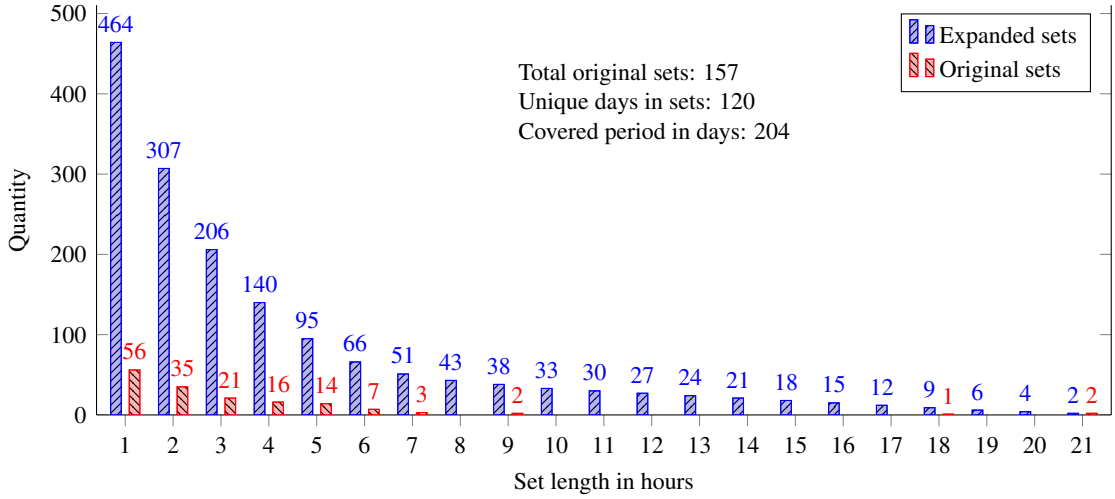


Figure 4.12: *Dataset expansions showing original and expanded unheated vacancy sets from the Golden dataset*

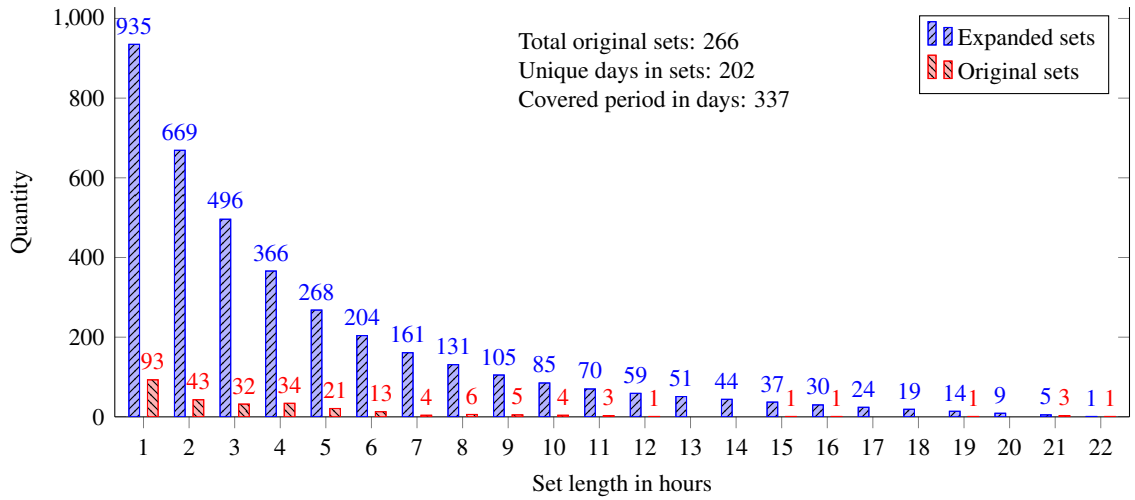


Figure 4.13: *Dataset expansions showing original and expanded unheated vacancy sets from the After Experiment dataset*

4.6 Selection of sensors used in modelling

A prediction model had two groups of variables. The first group were the predictors or independent variables: one or multiple sensors used as inputs into a model. The second group were the targets or dependant variables: data that was going to be the predicted output, dependant on the predictors in the model. The modelling methods were applied to find a fitting solution for the outputs when presented with the inputs.

For the output variable, the first floor landing air temperature was preferred. The reasons for selecting this sensor were as follows:

- The location of the sensor was in the 'middle' of the residential home which had three floors.
- The sensor was encapsulated inside a standard thermostat plastic enclosure, shielded from heat radiation. Furthermore the sensor was not near windows and therefore not subject to any sunlight.
- It was selected because it acted quickly on temperature changes induced by the heating system.

For the input variables, the following sensors were considered:

- First floor landing air temperature: Air temperature was used as input to reflect the current air temperature inside the residential home. It was taken into consideration because it represented the current state of the predicted sensor.
- First floor core temperature: The sensor was embedded into a brick wall and named 'core' because the wall was not connected to the outside. The core temperature was used because it represented the temperature of the residential home's thermal mass.
- Chimet outside air temperature and barometric pressure: The outside influences of current air temperature and pressures were selected to represent current weather conditions, supplied by a 3rd party. The sensor was 11 km away from the residential home to ensure that similar weather conditions were recorded.
- Calculated remaining day and night hours: A parameter indicating length of day and night. This aggregated sensor was selected because it indicated seasonal changes and daily cycles.
- Forecasted outside air temperature trends for up to 12 hours and for up to four days: The knowledge of future weather conditions was included to improve the prediction performance, enabling prediction models to account for outside changes and not just current weather conditions.

The sensors selected as inputs for the prediction model were chosen to reflect disparate sources of information, which were: website weather forecasts, outside current weather data and on-site based data collection.

4.7 Conclusion

In this Chapter the pre-processing of the raw data from disparate sources was discussed. The different sources were text files and an SQL database. The sensor data was imported into Matlab for further processing. The pre-processing applied data mining methods to prepare the raw data for further analysis and an objective exploration. The output of the pre-processing steps was a structure of sensors that held data as columns of timestamps and values for each sensor. The data was transformed to a normalized range and segmented into datasets for unheated vacancies.

The next Chapter uses the prepared data to create prediction models for the landing air temperature.

Chapter 5 Initial models

This Chapter introduces the creation of an initial model and the validation and improvements of inside air temperature prediction models in a vacant residential home. The initial models were created with historical environmental data from disparate sources.

Methods of model creation are presented followed by quality measures and the initial findings. The initial models were improved in several iterations and a prototype model is presented at the end of the Chapter.

5.1 Methods used

This Section introduces the methods of model creation and the model quality measure application to data that were considered for the research.

5.1.1 Linear regression models

The least squares method selected a linear model that fitted data based on the smallest sum of residuals. Residuals were the difference between actual and predicted value: $res = y - \hat{y}$. A basic linear regression model could be represented with a gradient b and a constant intercept term a as follows:

$$y(x) = a + bx \tag{5.1}$$

This regression model could be applied to data with one input x and one output y. The models described in this Chapter had multiple inputs and one output. In the literature inputs were referred to as independent variables or predictors and outputs as dependant variable. In a simple linear regression the coefficient b was calculated to meet the condition of having the smallest error compared to the original observed data.

In linear regression, several base models were defined. The models used in this Dissertation were linear, interaction, quadratic and pure quadratic regression models. The base formula for each is shown below. The terms were as follows: n was the number of inputs used in the model, b_i were the coefficients, x_i were the inputs, and $y(x)$ was the output. The term a was defined as '1', to make b_1 the constant intercept with the y-axis.

A linear regression model included all the inputs and a constant intercept term with coefficients for each as follows:

$$y(x) = b_1 a + b_2 x_1 + b_3 x_2 + b_4 x_3 + \cdots + b_{n+1} x_n \tag{5.2}$$

An interaction regression model consisted of linear terms, interaction terms and a constant

intercept term. Interaction terms were defined as the product of two inputs and one coefficient. The interaction regression model was as follows:

$$y(x) = \overbrace{b_1 a}^{\text{constant}} + \overbrace{b_2 x_1 + b_3 x_2 + b_4 x_3 + \cdots + b_{n+1} x_n}^{\text{linear}} + \underbrace{b_{n+2} x_1 x_2 + b_{n+3} x_1 x_2 + \cdots + b_{n(n+1)/2+1} x_{n-1} x_n}_{\text{interaction}} \quad (5.3)$$

A quadratic regression model included quadratic terms, linear terms, interaction terms and a constant intercept term:

$$y(x) = \overbrace{b_1 a}^{\text{constant}} + \overbrace{b_2 x_1 + b_3 x_2 + b_4 x_3 + \cdots + b_{n+1} x_n}^{\text{linear}} + \overbrace{b_{n+2} x_1 x_2 + b_{n+3} x_1 x_2 + \cdots + b_{n(n+1)/2+1} x_{n-1} x_n}^{\text{interaction}} + \underbrace{b_{n(n+1)/2+2} x_1^2 + b_{n(n+1)/2+3} x_2^2 + \cdots + b_{(n+1)(n+2)/2} x_n^2}_{\text{quadratic}} \quad (5.4)$$

A pure quadratic regression model consisted of linear terms, quadratic terms and a constant intercept term:

$$y(x) = \overbrace{b_1 a}^{\text{constant}} + \overbrace{b_2 x_1 + b_3 x_2 + b_4 x_3 + \cdots + b_{n+1} x_n}^{\text{linear}} + \overbrace{b_{n+2} x_1^2 + b_{n+3} x_2^2 + \cdots + b_{2n+1} x_n^2}^{\text{quadratic}} \quad (5.5)$$

5.1.2 Non-linear models

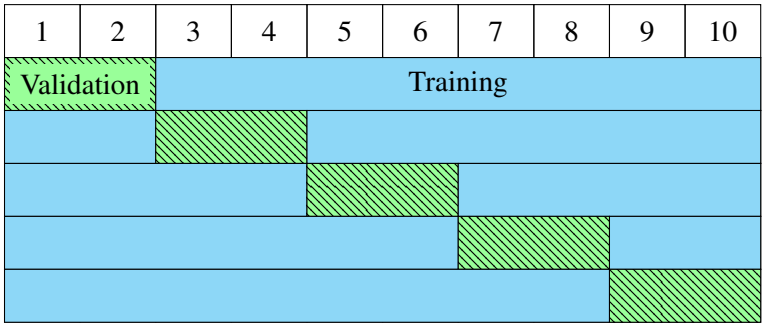
Non-linear models were an alternative method to linear models. Non-linear models were used in this work to create prediction models with Artificial Neural Networks (ANNs) in Matlab. A non-linear regression required a predefined function, which described a relationship between the inputs and the output. Basic relationships, such as exponential functions could be found by analysing plots with a logarithmic scale, showing a linear trend. ANN did not require prior knowledge and the training of a ANN created a prediction model with the desired inputs. A neural network was defined by nodes and their weights, which were adjusted towards a previously defined condition, for example a minimal error rate. The ANN was trained with a maximum number of iterations or a criterion, such as an acceptable error value. ANNs were used when an automation of finding relations between inputs and output was desired. Each successive creation of an ANN model with the same training inputs resulted in a unique ANN.

5.1.3 Application of the performance measures and cross validation

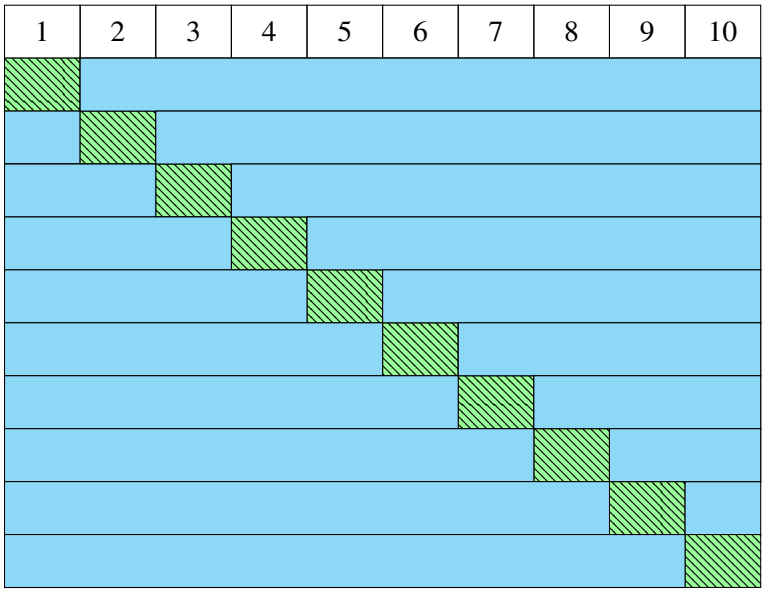
The prediction performance measures introduced in Section 2.6 on page 18 were applied to validation data and outputs of prediction models. The original recorded data was split into validation and training data to create unbiased performance measures for models, when applied to validation data.

A different method to test models was the application of cross validation. When cross validation

was used, all recorded historical data was used for training and validation. A selected percentage of observations from the original data was withheld to be used as a validation set and the remainder was used for model training. This step was repeated, until all recorded data was used for validation. For example: a value of 20% was selected for cross validation, which selected from the data the first 20% for validation and the remaining 80% for training. In a second step the second 20% were used for validation and the remaining 20%+60% were used for training. This was repeated 5 times in total and can be seen in fig. 5.1a. A variation of cross validation was Leave-One-Out Cross Validation (LOOCV), which withheld one dataset as validation set and the remaining data was used for training. An example diagram of the LOOCV can be seen in fig. 5.1b. The use of LOOCV made it possible to compare models with different inputs, as the performance results with LOOCV only changed when the inputs or the output were changed when the same data was used. The LOOCV was unbiased, because all data was used for training and validation, leaving no data unused in the validation and training process.



(a) *Cross validation with 10 datasets and a setting of 20%*



(b) *LOOCV with 10 datasets*

Figure 5.1: *Cross validation and LOOCV example datasets. The columns represent the dataset and the rows the iteration of the cross validation. The datasets for each iteration are marked as training and validation datasets.*

5.2 Dataset and model criteria

Before models were created, the data to be used for model creation was defined. The initial condition, set out in the introduction of this Dissertation was, that a vacant space was subjected to fewer dynamics affecting transient temperature developments, compared to inhabited space. The only dynamics to consider were the heating system and outside influences such as outside air temperatures, solar radiation, and rain. The second condition was, to select unheated space, which removed the heating system from the influences. The residential home vacancy was defined as valid, if it was at least one hour long. This minimum length was selected to ensure that there was a sufficient amount of time for transient temperature developments to be observed and that more than one data point from the 1-wire sensors were recorded. See Section 3.5 on page 30 and following for the retrieval intervals of each connected sensor.

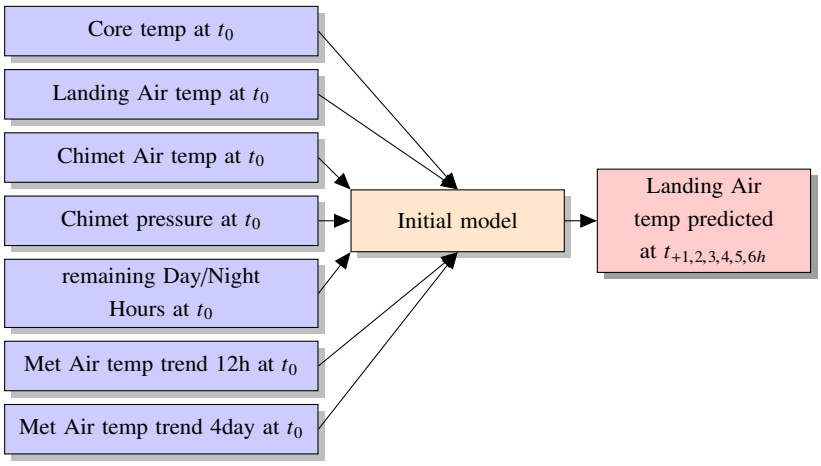
These conditions therefore constituted 'unheated vacant space' for the datasets considered in the initial model creation processes.

In a following step sensors used as inputs and one output were defined, as presented in Section 4.6 on page 57. *S13_FirstFloorLandingAir* was selected as the output variable for predictions. The input variables are listed in table 5.1. Block diagrams describing the initial model layout can be seen

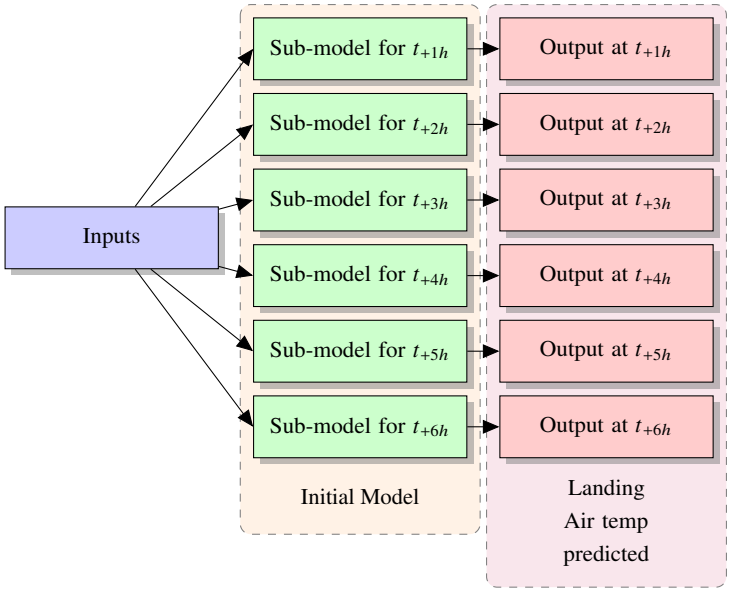
Table 5.1: *Sensors selected for the initial model*

S11_FirstFloorCoreTemp
S13_FirstFloorLandingAir
ChimetOutput_temp_C
ChimetOutput_pressure_mbar
remainingDayNightHours
metOutput_temp_C_trend_12hour
metOutput_temp_C_trend_4day

in figs. 5.2a and 5.2b. This model comprised of six sub-models, predicting 1, 2, 3, 4, 5, and 6 hour landing air temperature. t_0 was the starting point from which unheated vacant space predictions were made. t_{+xh} was a prediction that was x hours into the future, starting from t_0.



(a) *Initial model diagram*



***(b)** Sub-models and outputs in detail*

***Figure 5.2:** Model and sub-model block diagram*

The selection of training and validation sets was random. The training set was initially 70%, with 30% for validation sets. As vacancies were not evenly spread out throughout the week, each weekday was split into a 70% training and 30% validation datasets to ensure that data was selected randomly from all days of the week. For example, if 20 original datasets were found for Monday and 10 for Tuesday, then 14 training sets were randomly selected from Monday and 7 from Tuesday. The remaining 9 sets were then used as validation data. The frequency diagrams in fig. 5.3 show vacancies of different duration. The vacancies were not evenly spread across all weekdays, due to inhabitants' patterns.

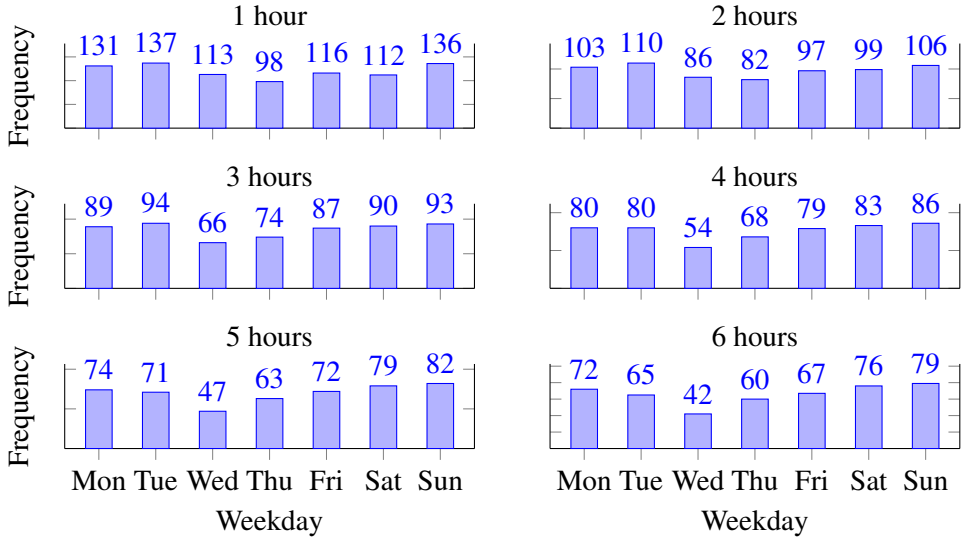


***Figure 5.3:** Expanded dataset distribution per hourly model per weekday*

5.3 Initial models

The initial models were created with the application of linear, interaction, quadratic and pure quadratic regression models. The coefficients b_x for the linear regression models were calculated with the goal of minimum prediction error for a given training set. The prediction sub-models returned a single prediction output for each prediction hour. For example: a 6 hour prediction model would consist of a 1, 2, 3, 4, 5, and 6 hour sub-model, predicting 1, 2, 3, 4, 5 and 6 hours. The initial models were set to predict up to six hours, based on the initial data explorations that showed a smaller number of individual sets existed for vacancy lengths longer than six hours, see Chapter 4 and fig. 4.12 on page 56. The training and validation sets created from the Golden dataset were used for each model, to ensure that comparisons would be based on the same data.

The linear regression model was represented by the following formula, which shows the coefficients b_x multiplied by the respective inputs:

$$\begin{aligned} S13\_FirstFloorLandingAir\_predict_{\text{linear}} &= b_1 + b_2\ S11\_FirstFloorCoreTemp \\ &+ b_3\ S13\_FirstFloorLandingAir + b_4\ ChimetOutput\_temp\_C \\ &+ b_5\ ChimetOutput\_pressure\_mbar + b_6\ remainingDayNightHours \\ &+ b_7\ metOutput\_temp\_C\_trend\_4day + b_8\ metOutput\_temp\_C\_trend\_12hour \end{aligned} \tag{5.6}$$

The model creation process took into account the 70/30 split for training and validation sets from the Golden dataset. The coefficients found for each of the sub-models in the linear regression model are given in table 5.2. The efficiency measures for the validation datasets are given in table 5.3. A simplified block diagram can be seen in fig. 5.4.

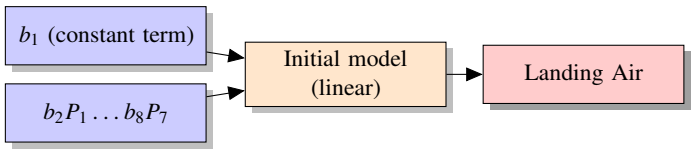


Figure 5.4: *Linear regression model layout*

Table 5.2: *Model parameters for a linear regression model*

	Prediction sub-model					
Coeff.	1 hour	2 hour	3 hour	4 hour	5 hour	6 hour
b_1	−0.032	−0.062	−0.083	−0.100	−0.119	−0.134
b_2	−0.204	−0.352	−0.474	−0.553	−0.609	−0.645
b_3	1.215	1.370	1.490	1.566	1.618	1.648
b_4	0.030	0.056	0.085	0.118	0.153	0.186
b_5	0.007	0.013	0.010	0.000	−0.004	−0.009
b_6	0.005	0.016	0.028	0.041	0.052	0.061
b_7	−0.004	−0.011	−0.014	−0.024	−0.028	−0.035
b_8	−0.002	0.000	−0.003	−0.009	−0.022	−0.038

These initial results showed, that a linear regression model was predicting better than a naïve forecast, based on the Mean Absolute Scaled Error (MASE) values being smaller than one. The predictive quality, measured by R^2, was consistently close to '1', indicating a qualitatively good fit.

Table 5.3: *Linear regression model efficiency measures calculated with validation data*

Hour	R^2	MAE/°C	MSE/°C^2	RAE	RSE	MASE
1	0.998	0.004	0.0000	0.033	0.002	0.727
2	0.997	0.006	0.0001	0.045	0.003	0.560
3	0.995	0.008	0.0001	0.060	0.005	0.523
4	0.993	0.009	0.0001	0.072	0.007	0.492
5	0.990	0.011	0.0002	0.082	0.010	0.466
6	0.987	0.013	0.0003	0.097	0.013	0.472

The coefficients calculated for the model formula were interpreted as follows:

- The coefficients showed consistent signs for each sub-model.
- The coefficients b_2 and b_3 had the highest values, relating to *S11_FirstFloorCoreTemp* and *S13_FirstFloorLandingAir*.

The results obtained from this initial model demonstrated that the model was able to predict temperature values. The model coefficients also identified influential inputs.

The second regression model was the interaction model. The first set of coefficients and inputs were the same as in the linear model shown in eq. (5.6). The additional terms were the 'interaction' terms, consisting of two inputs and one coefficient. The formula that described this model can be seen in eq. (5.7) and a block diagram representing this model can be seen in fig. 5.5.

$$\begin{aligned}
S13_FirstFloorLandingAir_predict_{\text{interact}} &= S13_FirstFloorLandingAir_predict_{\text{linear}} \\
&+ S11_FirstFloorCoreTemp \times (b_9\ S13_FirstFloorLandingAir \\
&+ b_{10}\ ChimetOutput_temp_C + b_{11}\ ChimetOutput_pressure_mbar \\
&+ b_{12}\ remainingDayNightHours + b_{13}\ metOutput_temp_C_trend_4day \\
&+ b_{14}\ metOutput_temp_C_trend_12hour) \\
&+ S13_FirstFloorLandingAir \times (b_{15}\ ChimetOutput_temp_C \\
&+ b_{16}\ ChimetOutput_pressure_mbar + b_{17}\ remainingDayNightHours \\
&+ b_{18}\ metOutput_temp_C_trend_4day + b_{19}\ metOutput_temp_C_trend_12hour) \\
&+ ChimetOutput_temp_C \times (b_{20}\ ChimetOutput_pressure_mbar \\
&+ b_{21}\ remainingDayNightHours + b_{22}\ metOutput_temp_C_trend_4day \\
&+ b_{23}\ metOutput_temp_C_trend_12hour) \\
&+ \ ChimetOutput_pressure_mbar \times (b_{24}\ remainingDayNightHours \\
&+ b_{25}\ metOutput_temp_C_trend_4day + b_{26}\ metOutput_temp_C_trend_12hour) \\
&+ remainingDayNightHours \times (b_{27}\ metOutput_temp_C_trend_4day \\
&+ b_{28}\ metOutput_temp_C_trend_12hour) \\
&+ b_{29}\ metOutput_temp_C_trend_4day \times\ metOutput_temp_C_trend_12hour
\end{aligned} \tag{5.7}$$

Using the training data, the resulting coefficients can be seen in table 5.4 with the efficiency measures obtained from the validation sets in table 5.5.

The results from the interaction regression model were similar to the results for the linear regression model: the quality measure R^2 was close to '1', and the model continually performed better than a naïve forecast, based on the MASE values. The quality measures regarding fitting

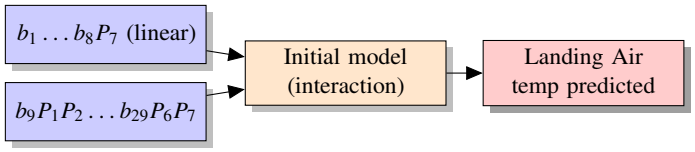


Figure 5.5: *Interaction regression model layout*

Table 5.4: *Model parameters for an interaction regression model*

	Prediction sub-model					
Coeff.	1 hour	2 hour	3 hour	4 hour	5 hour	6 hour
b_1	−0.057	−0.123	−0.193	−0.241	−0.306	−0.338
b_2	−0.016	−0.614	−0.882	−1.825	−3.225	−4.709
b_3	1.048	1.722	2.007	3.020	4.422	5.897
b_4	0.068	0.133	0.295	0.533	0.760	1.004
b_5	0.059	0.050	0.039	−0.054	−0.037	−0.089
b_6	−0.022	−0.040	−0.016	−0.020	0.016	0.040
b_7	−0.004	0.024	0.026	0.035	0.044	0.013
b_8	−0.007	0.062	0.045	−0.048	−0.165	−0.283
b_9	−0.065	−0.123	−0.204	−0.230	−0.259	−0.250
b_{10}	0.207	0.291	0.802	1.519	2.478	3.500
b_{11}	−0.206	−0.403	−0.329	−0.617	0.304	1.603
b_{12}	−0.232	−0.063	−0.211	0.077	0.187	0.337
b_{13}	0.345	0.629	0.795	1.514	1.533	1.399
b_{14}	−0.326	0.222	−0.132	−0.077	−0.072	−0.412
b_{15}	−0.178	−0.298	−0.757	−1.589	−2.547	−3.506
b_{16}	0.260	0.595	0.593	1.005	0.126	−1.243
b_{17}	0.245	0.060	0.174	−0.163	−0.323	−0.535
b_{18}	−0.414	−0.750	−0.917	−1.681	−1.657	−1.426
b_{19}	0.354	−0.260	0.082	0.035	0.033	0.334
b_{20}	−0.082	−0.188	−0.357	−0.566	−0.818	−1.087
b_{21}	0.012	0.028	0.031	0.070	0.111	0.130
b_{22}	−0.029	0.000	−0.095	−0.187	−0.320	−0.461
b_{23}	0.017	0.074	0.076	0.092	0.074	0.010
b_{24}	−0.011	0.066	0.086	0.161	0.166	0.253
b_{25}	0.070	0.086	0.191	0.309	0.421	0.568
b_{26}	−0.100	−0.193	−0.152	−0.034	0.107	0.396
b_{27}	−0.012	−0.064	−0.083	−0.133	−0.218	−0.306
b_{28}	0.055	0.047	0.048	0.021	0.039	0.016
b_{29}	0.023	−0.003	−0.006	0.012	0.041	0.056

errors, Mean Absolute Error (MAE), Mean Squared Error (MSE), Relative Absolute Error (RAE) and Relative Squared Error (RSE), all indicated a qualitatively good fit.

The coefficients b_x indicated two influential inputs, namely *S11_FirstFloorCoreTemp* and *S13_FirstFloorLandingAir*. Interaction coefficients with high values were: b_{10}, b_{13}, b_{15}, and b_{18}, relating to interactions of *S11_FirstFloorCoreTemp* and *S13_FirstFloorLandingAir* with *ChimetOutput_temp_C* and *metOutput_temp_C_trend_4day*. Their relationships did not have a significant influence in the short hour sub-models, but an influence was visible in the longer hour sub-models, judging by the coefficients' values.

Table 5.5: *Interaction regression model efficiency calculated with validation data*

Hour	R^2	MAE/°C	MSE/°C^2	RAE	RSE	MASE
1	0.998	0.004	0.0000	0.034	0.002	0.742
2	0.997	0.006	0.0001	0.047	0.003	0.585
3	0.995	0.008	0.0001	0.061	0.005	0.540
4	0.993	0.010	0.0001	0.075	0.007	0.514
5	0.990	0.011	0.0002	0.086	0.010	0.491
6	0.988	0.013	0.0003	0.100	0.012	0.484

The third model was created with a quadratic regression model. The quadratic model consisted of the interaction model and quadratic terms for each input. The resulting formula can be seen in eq. (5.8) and a block diagram depicting this model in fig. 5.6.

$$\begin{aligned} &S13\_FirstFloorLandingAir\_predict_{\text{quad}} = S13\_FirstFloorLandingAir\_predict_{\text{interact}} \\ &+b_{30}\ S11\_FirstFloorCoreTemp^2 + b_{31}\ S13\_FirstFloorLandingAir^2 \\ &+b_{32}\ ChimetOutput\_temp\_C^2 + b_{33}\ ChimetOutput\_pressure\_mbar^2 \\ &+b_{34}\ remainingDayNightHours^2 + b_{35}\ metOutput\_temp\_C\_trend\_4day^2 \\ &+b_{36}\ metOutput\_temp\_C\_trend\_12hour^2 \end{aligned} \tag{5.8}$$

The model's coefficients and efficiency measures can be seen in tables 5.6 and 5.7. As with the

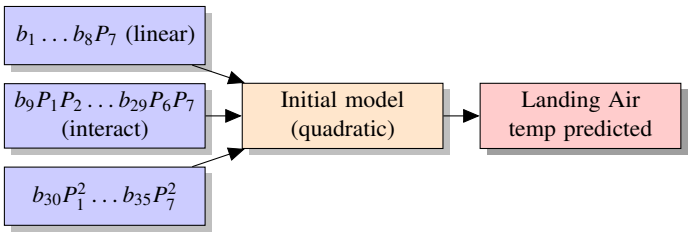


Figure 5.6: *Quadratic regression model*

first and second linear regression models, the quality measures showed the same tendency that the model performed better than a naïve forecast. The sub-models also showed a qualitatively good fit of the data according to R^2 and the fitting error measures. The coefficients b_{30} and b_{31} for the quadratic terms showed, that *S11_FirstFloorCoreTemp* and *S13_FirstFloorLandingAir* had a significant impact.

The model coefficients of the quadratic model were not all showing an increasing or decreasing tendency, as seen in the linear regression model, but showed diverse behaviour for the sub-models.

Table 5.6: *Model parameters for a quadratic regression model*

	Prediction sub-model					
Coeff.	1 hour	2 hour	3 hour	4 hour	5 hour	6 hour
b_1	−0.081	−0.157	−0.220	−0.334	−0.400	−0.497
b_2	−0.193	−0.758	−1.071	−1.349	−2.615	−3.381
b_3	1.237	1.886	2.231	2.470	3.697	4.291
b_4	0.053	0.050	0.213	0.559	0.916	1.370
b_5	0.081	0.113	0.025	0.039	−0.069	−0.078
b_6	0.077	0.116	0.171	0.153	0.155	0.080
b_7	−0.055	−0.043	−0.081	−0.021	0.099	0.268
b_8	0.013	0.118	0.161	0.098	−0.063	−0.203
b_9	−3.891	−0.270	0.831	−0.251	−7.025	−10.813
b_{10}	0.525	0.534	0.803	1.401	2.914	4.029
b_{11}	−0.265	−0.619	−0.504	−0.928	0.154	1.106
b_{12}	−0.096	0.103	−0.039	−0.004	0.075	0.084
b_{13}	0.391	0.746	0.997	1.647	1.295	0.806
b_{14}	−0.143	0.295	−0.062	−0.291	−0.187	−0.619
b_{15}	−0.560	−0.620	−0.801	−1.462	−3.011	−3.984
b_{16}	0.328	0.817	0.743	1.309	0.276	−0.644
b_{17}	0.104	−0.116	0.003	−0.058	−0.171	−0.219
b_{18}	−0.445	−0.853	−1.117	−1.773	−1.364	−0.756
b_{19}	0.149	−0.335	0.025	0.287	0.171	0.589
b_{20}	−0.062	−0.143	−0.254	−0.423	−0.707	−0.949
b_{21}	−0.020	0.007	−0.019	−0.009	0.018	0.035
b_{22}	0.000	0.055	−0.023	−0.187	−0.405	−0.686
b_{23}	0.021	0.074	0.045	0.034	0.024	−0.078
b_{24}	−0.003	0.060	0.104	0.172	0.206	0.315
b_{25}	0.050	0.069	0.218	0.273	0.350	0.438
b_{26}	−0.059	−0.155	−0.105	0.006	0.220	0.510
b_{27}	−0.018	−0.077	−0.118	−0.146	−0.209	−0.253
b_{28}	0.003	−0.046	−0.072	−0.102	−0.088	−0.071
b_{29}	0.032	0.011	−0.012	0.042	0.108	0.142
b_{30}	1.703	0.010	−0.444	−0.069	2.808	4.442
b_{31}	2.158	0.169	−0.587	0.110	4.036	6.201
b_{32}	0.026	0.052	0.029	−0.036	−0.067	−0.191
b_{33}	−0.044	−0.078	−0.069	−0.147	−0.068	−0.140
b_{34}	−0.045	−0.071	−0.079	−0.061	−0.055	−0.025
b_{35}	0.035	0.032	0.060	0.055	−0.013	−0.124
b_{36}	−0.013	−0.035	−0.064	−0.098	−0.118	−0.114

Table 5.7: *Quadratic regression model efficiency measures calculated with validation data*

Hour	R^2	MAE/°C	MSE/°C^2	RAE	RSE	MASE
1	0.999	0.004	0.0000	0.033	0.001	0.723
2	0.997	0.006	0.0001	0.046	0.003	0.573
3	0.995	0.008	0.0001	0.061	0.005	0.538
4	0.993	0.009	0.0001	0.073	0.007	0.506
5	0.991	0.011	0.0002	0.083	0.009	0.473
6	0.988	0.012	0.0002	0.094	0.012	0.459

The fourth model was the pure quadratic regression model. This model consisted of linear and quadratic terms only. The formula is shown in eq. (5.9) and a block diagram of this model can be seen in fig. 5.7.

$$\begin{aligned} S13\_FirstFloorLandingAir\_predict_{\text{pure quad}} &= S13\_FirstFloorLandingAir\_predict_{\text{linear}} \\ &+ b_9\ S11\_FirstFloorCoreTemp^2 + b_{10}\ S13\_FirstFloorLandingAir^2 \\ &+ b_{11}\ ChimetOutput\_temp\_C^2 + b_{12}\ ChimetOutput\_pressure\_mbar^2 \\ &+ b_{13}\ remainingDayNightHours^2 + b_{14}\ metOutput\_temp\_C\_trend\_4day^2 \\ &+ b_{15}\ metOutput\_temp\_C\_trend\_12hour^2 \end{aligned} \tag{5.9}$$

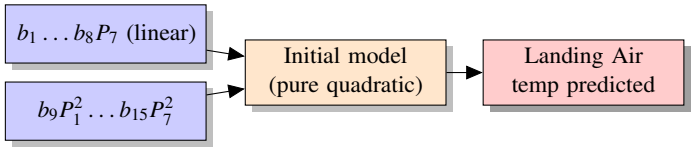


Figure 5.7: *Pure quadratic regression model layout*

The coefficients calculated with the training data can be seen in table 5.8 and efficiency measure results obtained from validation data can be seen in table 5.9. The pure quadratic regression model

Table 5.8: *Model parameters for a pure quadratic regression model*

	Prediction sub-model					
Coeff.	1 hour	2 hour	3 hour	4 hour	5 hour	6 hour
b_1	−0.058	−0.129	−0.175	−0.312	−0.420	−0.553
b_2	−0.073	−0.238	−0.366	−0.319	−0.247	−0.144
b_3	1.118	1.346	1.550	1.533	1.454	1.289
b_4	0.032	0.062	0.054	0.171	0.300	0.572
b_5	0.052	0.131	0.191	0.460	0.606	0.734
b_6	0.064	0.097	0.112	0.104	0.112	0.094
b_7	−0.047	−0.069	−0.087	−0.122	−0.117	−0.095
b_8	−0.003	0.013	0.017	0.023	0.024	0.007
b_9	−0.096	−0.086	−0.087	−0.209	−0.317	−0.425
b_{10}	0.066	0.006	−0.063	0.028	0.138	0.294
b_{11}	−0.003	−0.008	0.017	−0.041	−0.104	−0.260
b_{12}	−0.035	−0.092	−0.140	−0.354	−0.466	−0.565
b_{13}	−0.050	−0.071	−0.076	−0.056	−0.051	−0.027
b_{14}	0.044	0.057	0.070	0.097	0.087	0.053
b_{15}	−0.004	−0.022	−0.032	−0.043	−0.058	−0.052

Table 5.9: *Pure quadratic regression model efficiency measures calculated with validation data*

Hour	R^2	MAE/°C	MSE/°C^2	RAE	RSE	MASE
1	0.999	0.004	0.0000	0.033	0.001	0.728
2	0.997	0.006	0.0001	0.046	0.003	0.564
3	0.995	0.008	0.0001	0.063	0.005	0.557
4	0.993	0.010	0.0002	0.077	0.007	0.527
5	0.990	0.012	0.0002	0.089	0.010	0.508
6	0.987	0.014	0.0003	0.104	0.013	0.505

showed a better performance compared to the naïve forecast and qualitatively good results were observed for all performance measures. This model had one significant coefficient, b_3, which was related to *S13_FirstFloorLandingAir*.

5.3.1 Discussion

The results presented for the linear models indicated a qualitatively good fit to the validation data. In addition, the R^2 values decreased with increasing prediction length in longer time period sub-models. The MASE values however decreased with increasing lengths of time periods. MASE values reached a local minimum for the six hour sub-model. MAE, MSE, RAE, and RSE values rose with an increasing prediction length, resulting in a decreasing prediction accuracy for the model.

Not all the models were investigated further. Only the two best performing models were used: the linear model and the quadratic model. Both R^2 and MASE showed better values for a quadratic model, whereas differences between interaction and pure quadratic were marginal.

In these initial models, it was seen that certain inputs had a higher impact on the prediction than others, based on their coefficient values b_x. These were *S11_FirstFloorCoreTemp*, *S13_FirstFloorLandingAir*, *ChimetOutput_temp_C* and *metOutput_temp_C_trend_4day*.

The performance measure results for the models were mixed, and to differentiate models, the values R^2 and MASE gave the best indications. MASE was used to benchmark against a known model and R^2 gave a statistical view on the overall model's accuracy.

In a new approach, cross validation sets for LOOCV were used instead of validation datasets to create a different validation, containing all of the data. This method was used instead, to circumvent randomized models. The 70/30 randomized method created different model coefficients each time a new model was created.

5.4 First improvements

The first improvement was the change from randomly selected sets to LOOCV, which used unbiased sets for training and validation. The application of LOOCV to linear models resulted in constant coefficients and quality measures.

The improvements were applied to linear and non-linear modelling methods to evaluate the effects on the quality measures of predictions.

5.4.1 Linear models

The linear models considered were the linear and quadratic regression models. The results obtained with the LOOCV can be seen in tables 5.10 and 5.11. The quality measures showed a better performance than a naïve forecast. Only the quality measures for these models are shown, because the coefficients stayed the same for each run, due to the fact that LOOCV was used to compare models.

Table 5.10: *Linear regression model efficiency measures calculated with LOOCV*

Hour	R^2	MAE/°C	MSE/°C^2	RAE	RSE	MASE
1	0.998	0.004	0.0000	0.042	0.002	0.739
2	0.996	0.007	0.0001	0.059	0.004	0.613
3	0.993	0.008	0.0001	0.073	0.007	0.556
4	0.990	0.010	0.0002	0.083	0.010	0.512
5	0.987	0.011	0.0002	0.093	0.013	0.476
6	0.984	0.013	0.0003	0.105	0.016	0.463

Table 5.11: *Quadratic linear regression model efficiency measures calculated with LOOCV*

Hour	R^2	MAE/°C	MSE/°C^2	RAE	RSE	MASE
1	0.998	0.004	0.0000	0.041	0.002	0.716
2	0.996	0.006	0.0001	0.057	0.004	0.588
3	0.994	0.008	0.0001	0.071	0.006	0.542
4	0.991	0.010	0.0002	0.081	0.009	0.499
5	0.989	0.011	0.0002	0.089	0.011	0.457
6	0.987	0.012	0.0003	0.097	0.013	0.429

5.4.2 Non-linear models

The improvements were tested and validated against non-linear models. A neural network was used to investigate if a better model could be found compared to a linear model. The ANN model weights or coefficients were not investigated, because weights and connections within the ANN would change even when LOOCV was applied. The model benchmarking results for sub-models considering up to six hours can be seen in table 5.12. The creation of ANN models together with

Table 5.12: *ANN model efficiency measures calculated with cross validation*

Hour	R^2	MAE/°C	MSE/°C^2	RAE	RSE	MASE
1	0.998	0.004	0.0000	0.039	0.002	0.024
2	0.997	0.006	0.0001	0.050	0.003	0.031
3	0.995	0.007	0.0001	0.060	0.005	0.038
4	0.994	0.008	0.0001	0.064	0.006	0.041
5	0.994	0.008	0.0001	0.066	0.006	0.043
6	0.992	0.009	0.0002	0.071	0.008	0.046

the application of LOOCV was computationally expensive, because the ANNs were trained and validated for each set. Each set was iterated through to achieve a required criterion, which was a MSE value of smaller or equal to 0.0001 °C^2. Due to the iterative creation of neural nets, calculation times between 10 s and 100 s per run for one set were observed. The total number of sets was 892 and calculations took 12 hours to complete.

5.4.3 Discussion

The lack of long periods of vacancy limited the number of hours used to create prediction models. The longest periods of vacancy with a sufficient number of original sets was six hours long. Data for longer periods of vacancy was available but not in sufficient numbers and frequency.

LOOCV was used to compare the overall performance of a model with the inputs and therefore was used to compare different models with different inputs or different modelling methods, in this case linear and non-linear models. The computational expense for LOOCV increased when calculations of all possible input combinations were carried out. It was impractical to accomplish in a reasonable amount of time, especially for ANNs. The reasoning behind choosing LOOCV, was the ability to compare models based on quality measures without bias.

The efficiency measures for the non-linear models were slightly better than for the linear models, but the differences were small. This led to the deduction that even a linear model was delivering an able fit to predict temperatures. ANNs were not further investigated in this Dissertation, because of their intensive computational needs and marginally improved prediction performance.

5.4.4 Issues with selected data

As pointed out in the discussion on first improvements and from the analysis on the amount of vacancy datasets, there was a discrepancy between existing long sets and the impact they had in terms of expanding the overall sets. For example a set as long as 20 h yielded 20 one hour long sets, 19 two hour long sets, 18 three hour long sets, et cetera. To address this issue, each expanded vacancy set was marked with a so-called 'unique identifier'. The restrictions that were applied to the datasets were as follows:

1. Limitation of acceptable duration of day-to-day vacancies regarding the maximum length. 24 h was set as maximum, above this threshold vacancy sets were seen as holiday time and therefore did not qualify to be taken as everyday occurrences of vacancy.

2. Marking the multiple subset creation from one set of vacancy data by assignment of unique IDs and dates. This limitation did not decrease the number of training or validation sets, but how they were selected: by unique ID, rather than by each individual set.

3. Change in the LOOCV approach to use unique IDs referring to a whole vacancy set. Each set was used as individual set and not expanded to create more hourly data. The Golden dataset had 157 unique sets, which were used instead of the expanded sets for each sub-model.

The changes were implemented and evaluated in a second improvement step to limit the data selection process and performance analysis.

5.5 Further improvements

The impact of vacancy dataset length limitation to a maximum of 24 hours resulted in the exclusion of four vacancy sets in the Golden dataset. The four long vacancy sets accounted for half of the available expanded datasets. The inclusion of these very long datasets could have negatively impacted the models by over-fitting.

The results for R^2 and MASE are shown in tables 5.13 and 5.14. All regression models were used and compared against each other.

Table 5.13: *Model efficiency measure R^2 calculated with cross validation*

Model \ Hour	1	2	3	4	5	6
linear	0.996	0.992	0.985	0.977	0.965	0.964
interaction	0.996	0.991	0.978	0.965	0.752	0.061
quadratic	0.996	0.990	0.971	0.910	0.224	0.460
pure quad.	0.997	0.993	0.987	0.980	0.936	0.892

Table 5.14: *Model efficiency measure MASE calculated with cross validation*

Model \ Hour	1	2	3	4	5	6
linear	0.897	0.854	0.873	0.860	0.760	0.520
interaction	0.957	0.868	0.998	1.108	2.010	2.469
quadratic	0.932	0.908	1.117	1.358	2.938	1.774
pure quad.	0.845	0.781	0.822	0.829	1.049	0.887

The MASE values showed that the sub-models or entire models were not better than a naïve forecast. Apart from the linear regression model, all other models were at least once above a value of '1' in one of the sub-models. The measure R^2 showed different results, marking all models to be suitable fits.

5.5.1 Discussion on further improvements

The second improvement minimized the datasets significantly. The efficiency measures dropped sharply for predictions of five and six hours. This was due to fewer sets being available, imposed by the limitations on allowed vacancy lengths and the use of unique sets rather than single expanded sets with LOOCV. The new distribution of the Golden dataset can be seen in fig. 5.8.

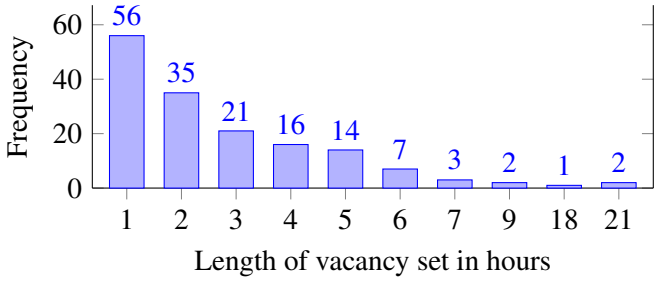


Figure 5.8: *Distribution of vacancy set lengths for the Golden dataset*

The MASE efficiency measure showed, that only the linear regression model was producing better predictions than the naïve forecast for all sub-models. The second best model was the pure quadratic one, showing only a weaker performance in the five hour sub-model. The remaining models, including the ANNs, were not able to perform better, also coinciding with their R^2 values being smaller than the two best regression models. Although the MASE results seemed to be 'inverted' compared to the R^2 values, its added benefit was having an additional quality measure. A

prediction model that would perform worse than a naïve forecast was not suitable for prediction application.

5.6 Considerations on model accuracy, sensors and predicted data

The initial linear prediction model performed generally better than a naïve forecast, according to the quality measures MASE and R^2. One question, which has not been pursued was the implications on the predicted data, in terms of temperature, the differences in the quality measures and the sensors' capabilities. The results retrieved by the models were compared with general sensor values and error ranges.

The Golden dataset was used to derive the models. In addition it was the base for the calculations for the quality measures with these models and the original data. The landing air temperature's statistical data can be seen in table 5.15. The table shows the overall data statistics and data for the temperature ranges encountered during a set and also the digital steps for the temperature sensor for the set ranges.

Table 5.15: *Absolute temperature values, ranges and absolute error statistics on landing air temperature and digital sensor steps from the Golden dataset and a linear regression model*

	All sets	Vacancy range		1st Model $\|err\|$	
Value name	(°C)	(°C)	Steps	(°C)	Steps
mode (most frequent)	21.5625	0.0625	1	0.0000	0
mean (average)	21.8327	0.1242	2	0.0875	2
median (middle value)	21.7500	0.0625	1	0.0660	2
minimum	19.1875	0.0000	0	0.0000	0
maximum	24.5625	0.6875	14	0.4141	7
range	5.3750	0.6875	14	0.4141	7

Comparing those values with the error margins and accuracy of the DS18B20 temperature sensors, the temperatures were within a few digital steps. The accuracy of the DS18B20 was ± 0.5 °C, the repeatability error margin was ± 0.2 °C, both provided by the datasheet. All sensors in use were from the same batch and were tested simultaneously, showing a maximum temperature spread of 0.125 °C between them. The resolution used was 12 bit, resulting in a sensitivity of 0.0625 °C for one digital step. Those values and the relation to the digital steps can be seen in table 5.16.

Table 5.16: *DS18B20 characteristics*

Characteristic	°C	Digital steps
Digital step	0.0625	1
Accuracy	± 0.5	± 8
Repeat accuracy	± 0.2	± 4
Calibration spread	0.0125	2
Maximum[1]	+85	-
Minimum	-10	-
Range	95	1520

5.6.1 Discussion

The quality measures R^2 and MASE gave an indication about which of the models was better, but did not give an indication of the actual errors between the predictions and the actual landing air temperature. By taking into account the specifications of the sensor used to record the data, it seemed that the most frequent error from the prediction, 0.0875 °C and 0.0953 °C, would not be distinguishable because of the repeat accuracy of ±0.2 °C.

Based on these findings, the initial models were within the error margins of the sensors used in this research and therefore the models were accurate for predictions of the landing air temperature.

The presented models were all based on the Golden dataset, which spanned from April 2011 until October 2011, making the model creation highly dependant on this particular set. The testing of different datasets was possible and the options were as follows:

- The use of data after the experiment finished, or
- expanding the datasets to include night time, regarding it as quasi-vacant space due to sleep of the inhabitants, or
- a combination of vacancy alone, vacancy and night time, night time only to create different sets.

Night time data was not the initially intended target, but it showed decreased dynamics in terms of inputs and the inhabitants' influence on the residential home - apart from body heat.

However, those additional datasets were taken into consideration carefully, as the structural layout of the residential home was changed after the experiment, which could have influenced temperature developments. The Golden dataset was considered as the 'gold standard' for comparisons and the additional datasets as estimations of how the model creation and comparison values would change, when different data was used.

The latest models were created with LOOCV for each dataset and the coefficients derived from averaging across all the created models. As all models were created with linear regression, these models did not produce different coefficients compared to a linear regression applied to the entirety of a dataset. Therefore LOOCV did not contribute towards the creation of a better or different model on this occasion. The coefficients could be derived by a linear regression instead of LOOCV and averaging.

Furthermore, due to the use of all datasets for training and validation at the same time, with LOOCV, the created models were the best possible, as all the validation data was known to the models. The knowledge of all validation datasets rendered the quality measures ineffective; validation measures were created by presenting unknown data to a model. The created models were used to differentiate between the models of a particular dataset, but not to validate models against unknown data.

[1] with accuracy of ± 0.5 °C

5.7 New validation method

The use of random data had shown that modelling results had repeatability problems, as model coefficients changed with each model creation process. Therefore, it was not possible to obtain the best model during this process.

The new proposed approach was to build a continuous system. This system would relate to a real world application: modelling and validation during data collection, a so-called learning system.

5.7.1 Iterative approach

The approach to create models for landing air prediction were based on historical sensor data. A new approach was considered, changing the method into an iterative approach that would simulate real world behaviour: creation of the models over time, going from day to day. This included creation and validation of data.

The partitioned datasets were used, and the method of model creation was changed as follows:

- Build model from 1 to n datasets
- Validate against dataset $n + 1$
- Save validation and quality measures for analysis
- Increase n by 1

The variable n was a number ranging from '1' to the maximum number of sets in a dataset.

The models used in this approach were the linear regression models, which showed the best results in Section 5.5 on page 72. The model that was used for the new validation method can be seen in fig. 5.4 on page 64.

Results for the Golden dataset can be seen in fig. 5.9. The number of datasets for validation is shown and a percentage of a better result than the naïve forecast on the y axis. This measure was generated with MASE values, converting values above and equal to '1' to '0' and values smaller than '1' to '1'. The calculation was carried out for every dataset, added up, and divided by the current total number of datasets, creating a new benchmarking measure %MASE<1. For example: a sub-model iterated through 100 datasets, which had 60 sets with MASE values smaller than '1', the %MASE<1 value would therefore be 60%. This measure was used to establish if a model would be suitable for predictions.

The graph in fig. 5.9 shows a decline when the datasets used for validation were worse than the naïve forecast. An increase showed an improved performance for the model. The graph shows a decline for longer hour sub-models, coinciding with a declining number of datasets available for long term vacancies. Furthermore all of the sub-models showed a declining performance with an increasing number of sets.

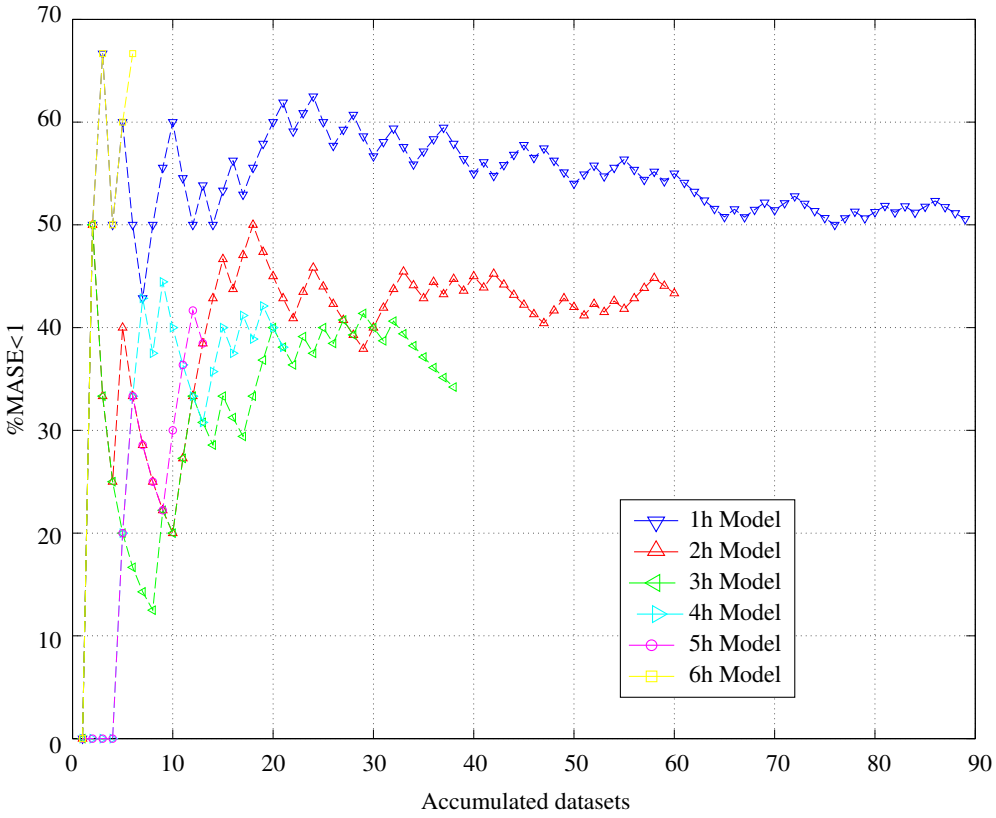


Figure 5.9: *Prediction better than naïve forecast over datasets used for validation showing each sub-model*

5.7.2 Reduction of sensor inputs

The choice of inputs to the models were based on the premise that an inclusion of sensors from disparate sources would improve the prediction rate. The old sensor list can be seen in table 5.1 on page 62.

Sensors were removed, as they were not believed to have a significant impact, or because they were not used in a suitable manner. The reasons for their removal and replacements were as follows:

- The trending of the *Met Office* forecasts with lengths of four days and one for 12 hours was based on the maximum forecast and not on the maximum vacancy length. These sensors were replaced by a dynamically changing input for each sub-model. That was, a one hour Met Office forecast for the outside air temperature for the one hour sub-model; two hour forecast for the two hour sub-model, et cetera, until the sixth hour forecast for the six hour sub-model.

- The *remainingDayNightHours* was introduced to capture the amount of day and night time. Temperature drops during the night and the introduction of the *Met Office* hourly forecasts gave feedback about night time or daytime via temperatures. Therefore the sensor was redundant and removed from the list of potential inputs to further models.

- The *Chimet* pressure was an instantaneous value and not trended, therefore not returning information about pressure changes over time. The sensor was removed and not replaced.

The new sensor inputs for the model are shown in table 5.17.

Table 5.17: *Sensors selected in new iterative model*

Sensor name
S11_FirstFloorCoreTemp
S13_FirstFloorLandingAir
ChimetOutput_temp_C
metOutput_temp_C_x_hour

The selected sensor inputs were then used to create a reduced sensor input iterative model and %MASE<1 values were calculated for the next day validation set in figs. 5.10 and 5.11. A simplified block diagram is shown in fig. 5.12. The Met Office forecast was applied to each sub-model with a different forecast period. The term t_{+xh} was used to indicate hourly forecasts of up to six hours. t_0 was the start of a prediction of unheated vacant space.

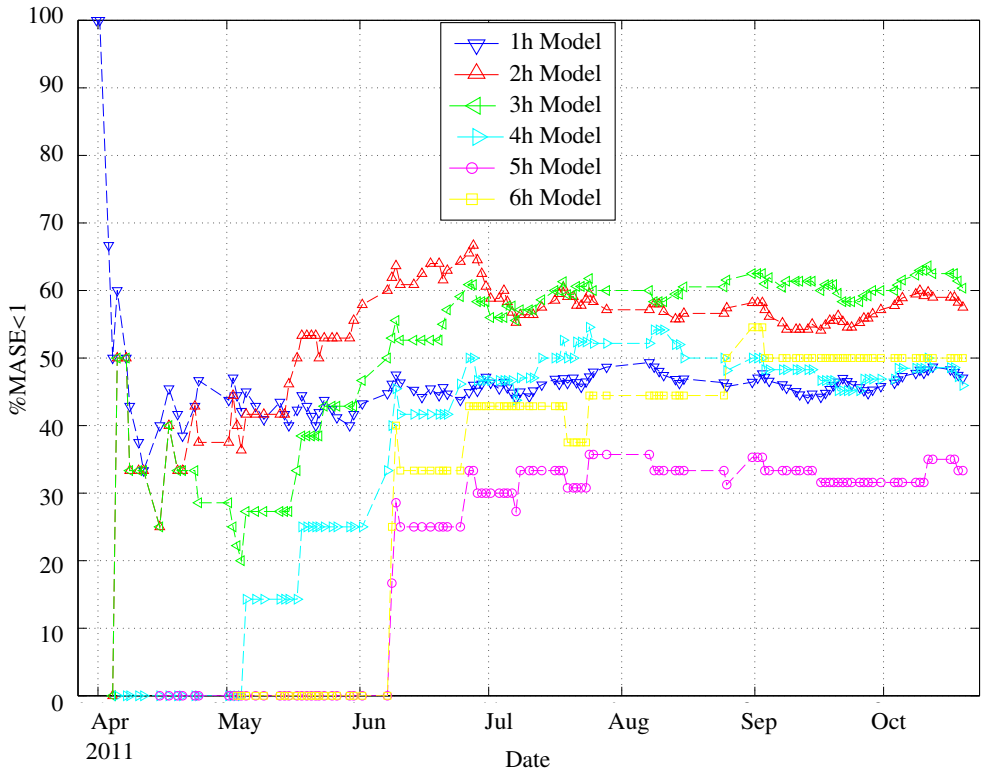


Figure 5.10: *Iterative model benchmark created with unheated vacancies from the Golden dataset*

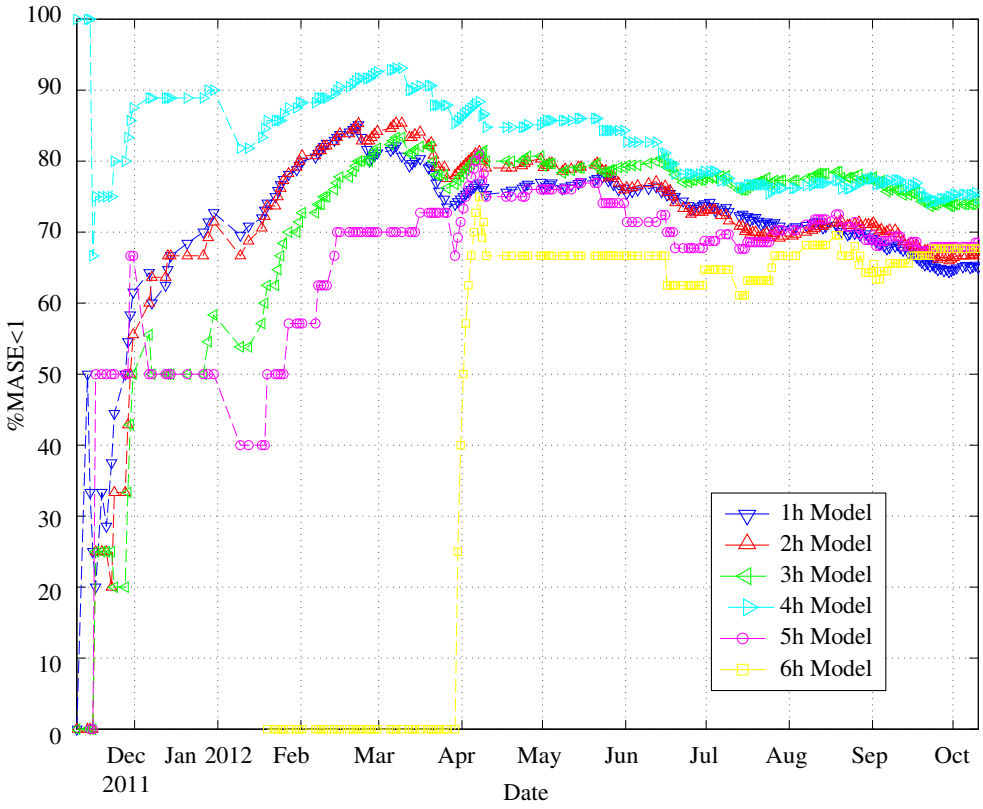


Figure 5.11: *Iterative model benchmark created with unheated vacancies from the After Experiment dataset*

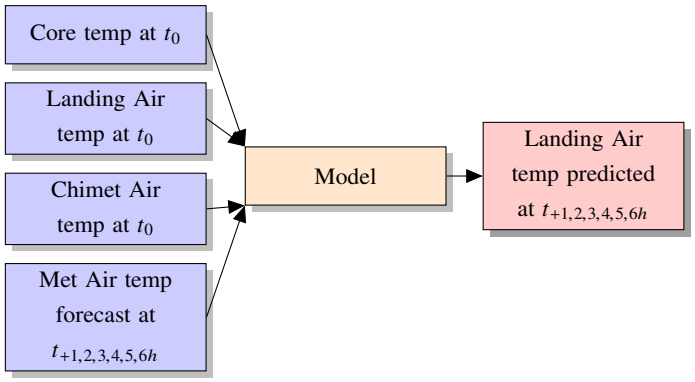


Figure 5.12: *Reduced sensors model layout*

The graphs show the predictions' performance development over time. A horizontal line was observed, when there were no further validations available for a set, an increase was observed when the MASE value was higher than 1 and a decrease was observed when a validation set had an MASE value smaller than 1.

The plots in figs. 5.10 and 5.11 show different datasets: Golden and After Experiment. Two particular phenomena were observed: the Golden dataset performed worse than the After Experiment and the After Experiment dataset showed a performance recovery towards the end. The slight increase in prediction success rate could have resulted from data that was recorded in a similar seasonal period before and therefore reinforced the training data for that particular period. The

plot in fig. 5.10 shows that models created from the Golden dataset had performance rates around 50%. The After Experiment dataset showed an improvement compared to the Golden dataset, with results in the range of 60-80%.

5.7.3 Limiting the magnitude of prediction output

The iterative models did not yield sufficient %MASE<1 values when applied to the Golden dataset unheated vacancies. In an additional step the predicted output was limited to improve the prediction quality of the models.

A statistical analysis of the temperature changes in unheated vacancies within one hour showed, that a temperature change of $0.0625\,°C\,h^{-1}$ was a common value. The iterative model was created with day-to-day data and therefore the output could be limited to an anticipated range. The statistical values of interest were as follows:

- mean: the average value.
- mode: the most frequent value.
- median: the middle value.

A simplified block diagram shown in fig. 5.13 depicts the integration of output range limitations for the iterative model.

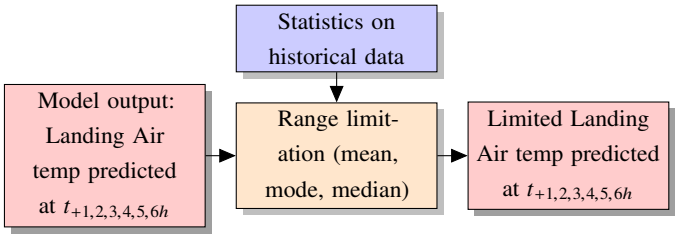
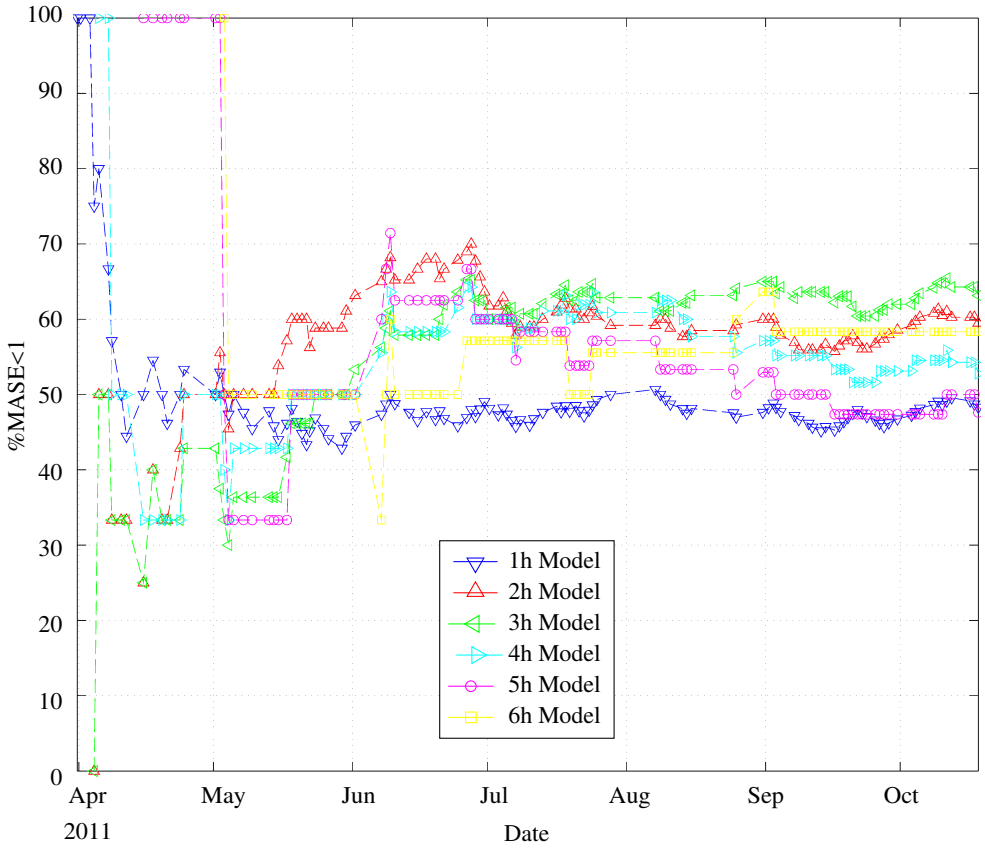
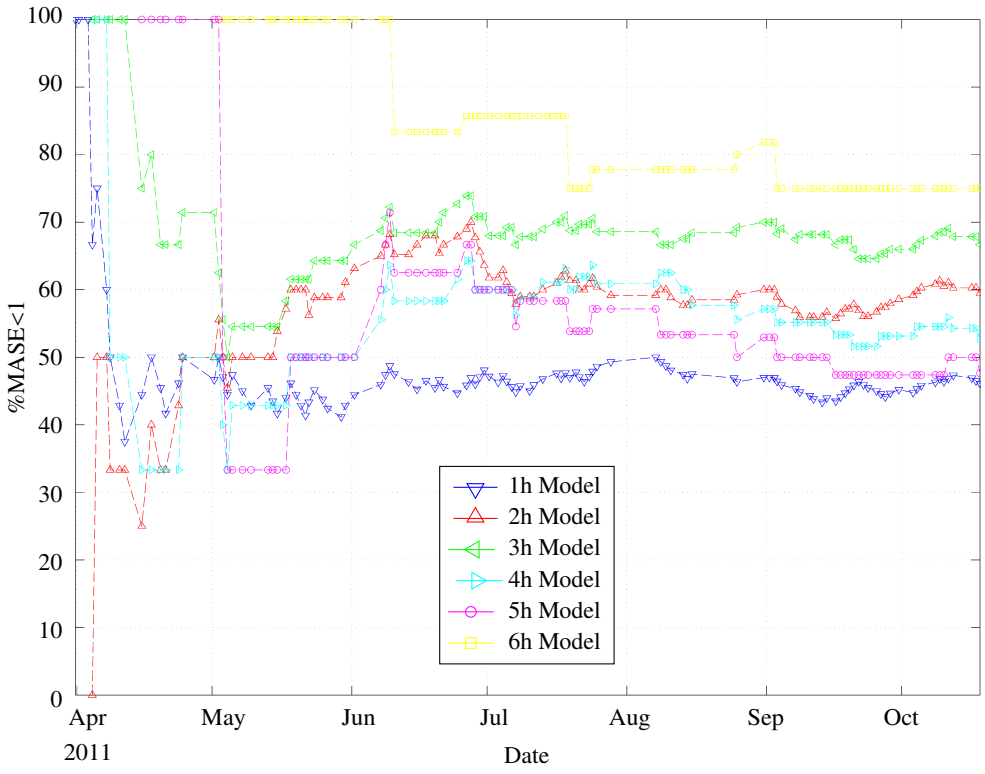


Figure 5.13: *Restricted model output with statistics derived from historical data*

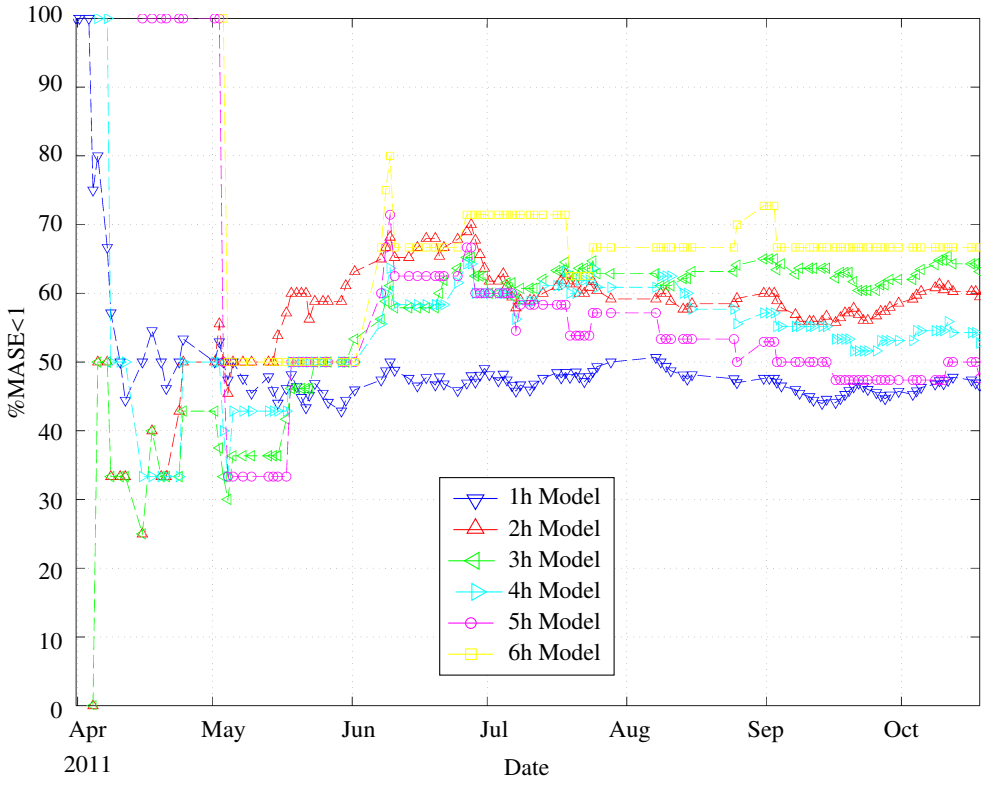
The statistical values were generated from the landing air sensor data during the iterative model training process. The plots in figs. 5.14a to 5.14c, show the %MASE<1 benchmarking results for the prediction output range limitation application for mean, mode, and median respectively.



(a) *Prediction benchmarking results with mean value prediction range restriction*



(b) *Prediction benchmarking results with mode value prediction range restriction*



(c) Prediction benchmarking results with median value prediction range restriction

Figure 5.14: *Prediction benchmarking with prediction output range limitation methods using unheated vacancies from the Golden dataset*

The impact of the prediction output range limitation can be seen in the longer hour sub-models, where the improvements strengthened the model performance in the early datasets. Longer time period sub-models had only a few datasets to train from and therefore the errors made for a prediction were higher compared to later predictions. The best performing method for all models was 'mode', which increased the performance of the longer hour sub-models over the 50% mark for the Golden dataset. The short hour sub-models, specifically one and two hours, were not significantly affected. This was attributed to the fact that those models had further data to learn from and therefore the predictions did not have high error rates, that would be affected by a prediction output range limitation rule.

5.7.4 Impact of selected sensors on the model outcome

The models used for predictions had a set of inputs selected from the sensors, based on the hypothesis that disparate sources would lead to an improved model. However it was not considered that this may not be the true for the selected sensors. Investigations into the analysis of combinations of the available sensors were carried out to address this issue.

Reinforcing the importance of past values for a temperature prediction were supported with two additional inputs: core temperature trending and landing air trending. Those two inputs reflected the temperature trend for 1-6 hours into the past, that was a linear fit onto the data points found within those hours and the gradient was used as the sensor value. These derived sensors were then

used as additional inputs.

With the reduced inputs and aforementioned new trending inputs, the following sensors were considered to be part of the model:

- Landing air as a permanent input, as this was also the output and selected because it had the highest influence, see Subsection 5.3.1 on page 70.
- All combinations of Chimet air temperature and core temperature, as those were the second and third most important inputs discussed in Subsection 5.3.1 on page 70.
- One of each trend, to include past values: core and landing air temperature and trends that use 1 to 6 hours of data.
- The Met Office forecast trends for 1 to 6 hours, applied to each according sub-model to include a forecast. For example, the one hour forecast trend was used in the one hour sub-model, the two hour forecast in the two hour sub-model, until the six hour forecast trend and six hour sub-model.

This yielded 392 possible models and the quality measures collected were MASE, %MASE<1, R^2, and the maximum error. Those were used in addition to a visual confirmation of plots to select the best models that either minimised errors, had a small error rate in general, the highest successful predictions or highest percentage of successful predictions over time. The Golden dataset and the After Experiment dataset were used as basis for the historical model creation process. The models did not take into account the prediction range restrictions introduced in Subsection 5.7.3 on page 80.

The block diagram in fig. 5.15 shows the new approach. The new variable was named t_{-xh}, stating that a temperature trend was generated from x hours before t_0.

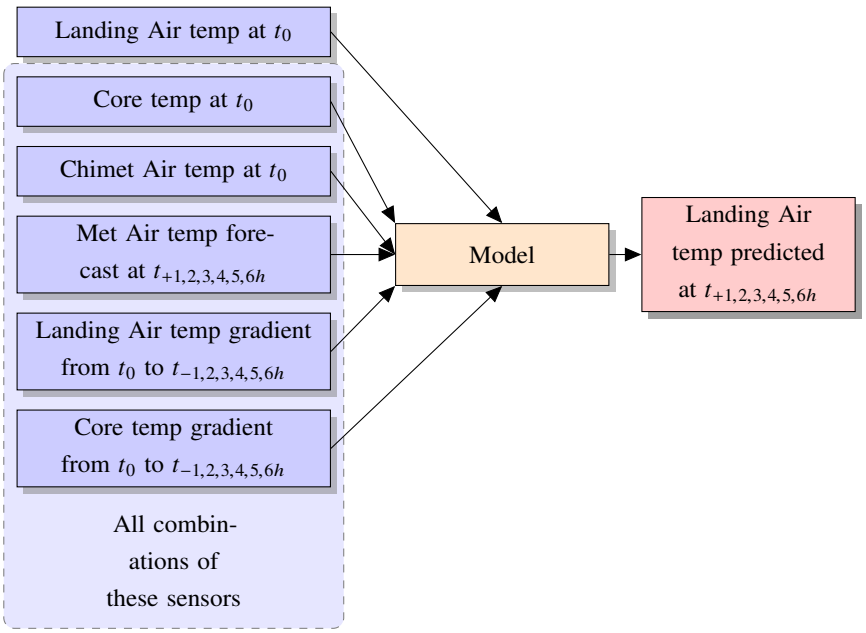


Figure 5.15: *New and old sensor inputs for selection of best model*

The results for the best sub models can be seen in fig. 5.16.

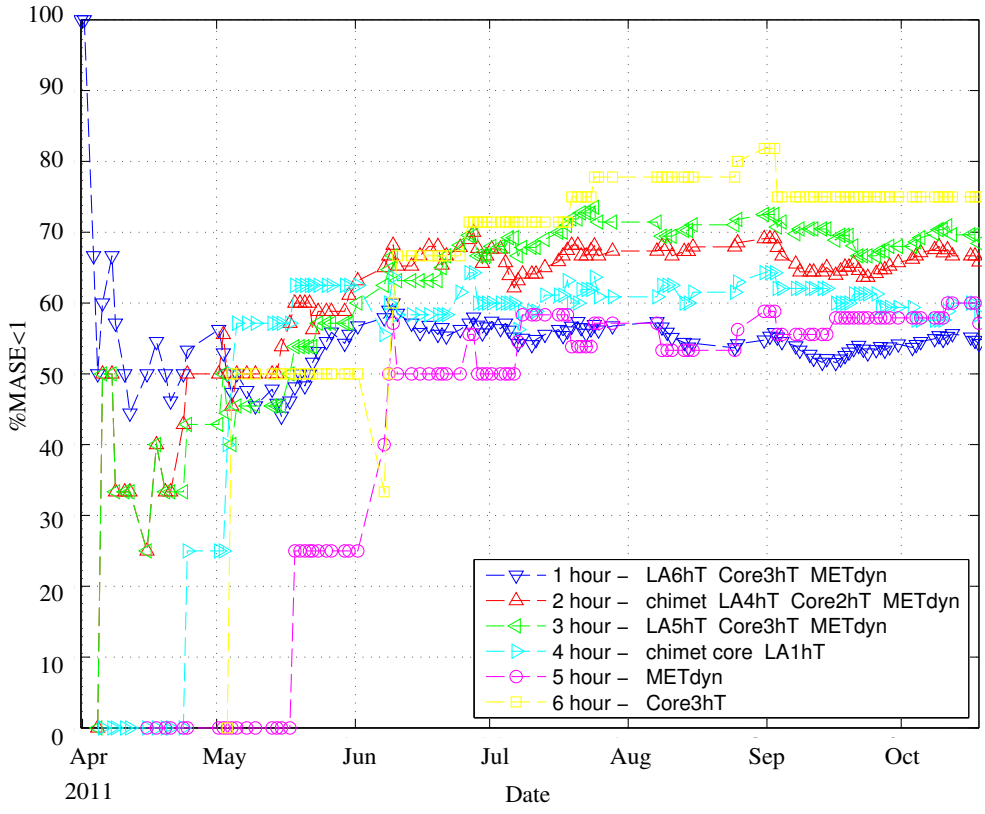


Figure 5.16: *Best sub-models selected by the highest percentage of being better than a naïve forecast*

As can be seen in the figure and compared with fig. 5.14a on page 81, the performance at the beginning of each model was worse, because no prediction output range limitation was applied to the models. Overall the models' performance ranged between 50% and 80%, which was an improvement. The final model block diagram, consisting of each individual sub-model, can be seen in figs. 5.17 to 5.22. Each *Sub-model x* corresponds to the results for *x hour* shown in fig. 5.16.

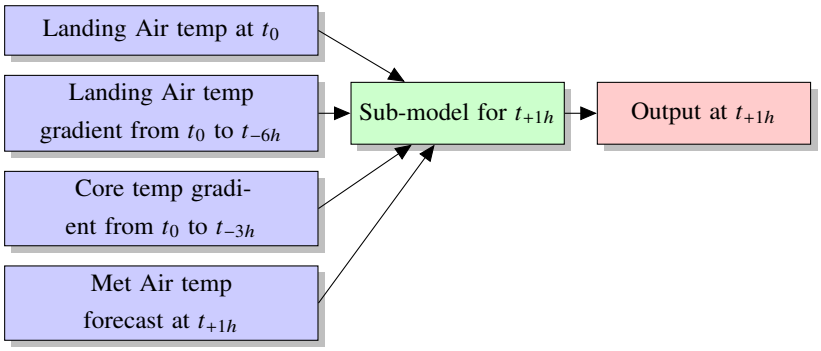


Figure 5.17: *Sub-model 1 block diagram*

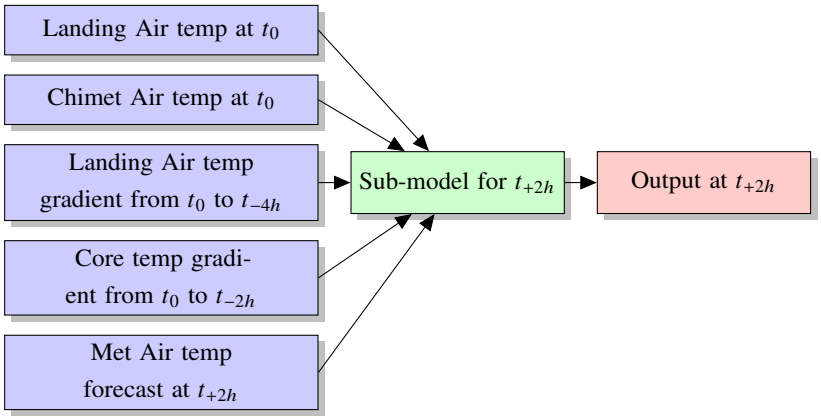


Figure 5.18: *Sub-model 2 block diagram*

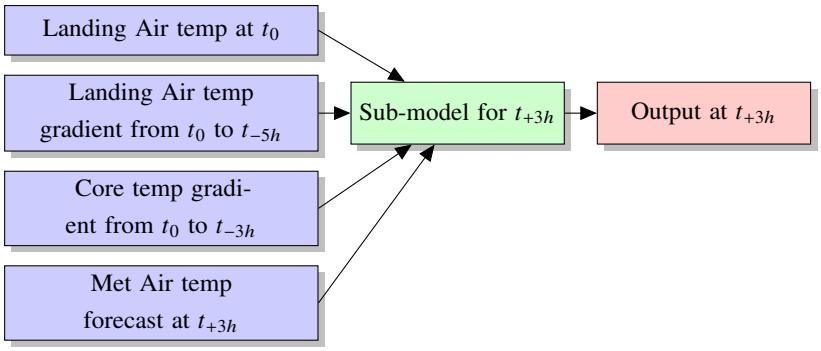


Figure 5.19: *Sub-model 3 block diagram*

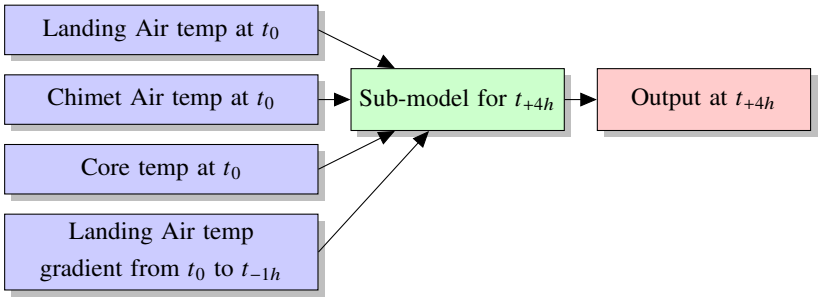


Figure 5.20: *Sub-model 4 block diagram*

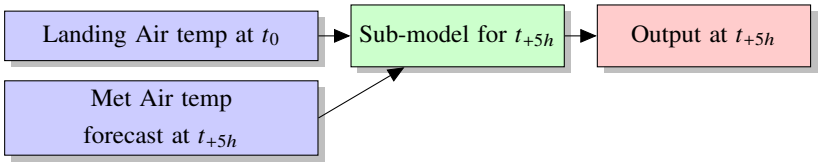


Figure 5.21: *Sub-model 5 block diagram*

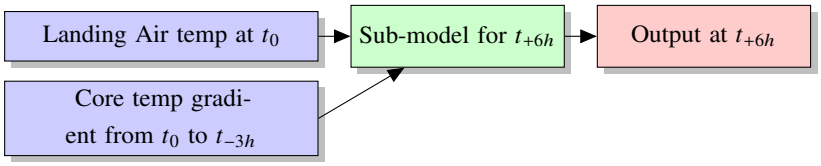


Figure 5.22: *Sub-model 6 block diagram*

The results obtained with this method showed, that different inputs for each sub-model improved the individual sub-model prediction performance. Furthermore, the results also showed that forecasts and temperature trends improved the sub-model performance.

5.8 Conclusion

In this Chapter initial models were built that were described, compared and enhanced to predict inside air temperature in a unheated vacant space in a residential home, which used data from disparate sources of information. The models used were linear regression models - linear, interaction, quadratic and pure quadratic - in addition to non-linear models - ANN. During improvement processes, the efficiency measures for prediction models were represented by R^2 and MASE, where R^2 described the model's ability to fit data and MASE benchmarked a model against a naïve forecast. The use of both measures enabled different views of the data, one on the quality overall and one on benchmarking against the naïve forecast.

The results showed that models with fewer parameters, such as linear and pure quadratic regression models, showed an increased prediction performance.

An investigation into the Golden dataset was undertaken to compare initial prediction results with the data found in the set. This revealed that the temperature measurements and predictions were within a range of 1 °C. It was concluded that models performed well in light of these small temperature ranges, which in some cases matched the measurement accuracy of the equipment.

A new validation method was proposed to address the problems with LOOCV and linear models: creating a model 'on-the-go', that was building up a model from day one of the historical dataset through to the final day, instead of all data at once. This made it possible to take a unique look at the model performance development over time. The prediction output range limitation, limiting the possible maximum and minimum output changes per hour, increased the overall model performance.

A new modelling approach helped identify inputs which created the best model. The method showed that for each sub-model, different inputs yielded a better prediction than a naïve forecast. The use of different relevant inputs increased the benchmarks overall, but the overall prediction benchmarks remained between 50-80%. These models were seen as a starting point for future research and not further investigated in this Dissertation.

Furthermore, the results showed that prediction lengths of longer than six hours did not yield reliable models. This was due to the inhabitants use patterns of the residential home, which caused only a small number of long vacancies.

This Chapter set the ground work for model creations and the efficiency measures, which were also used to compare and benchmark the models in Chapter 6.

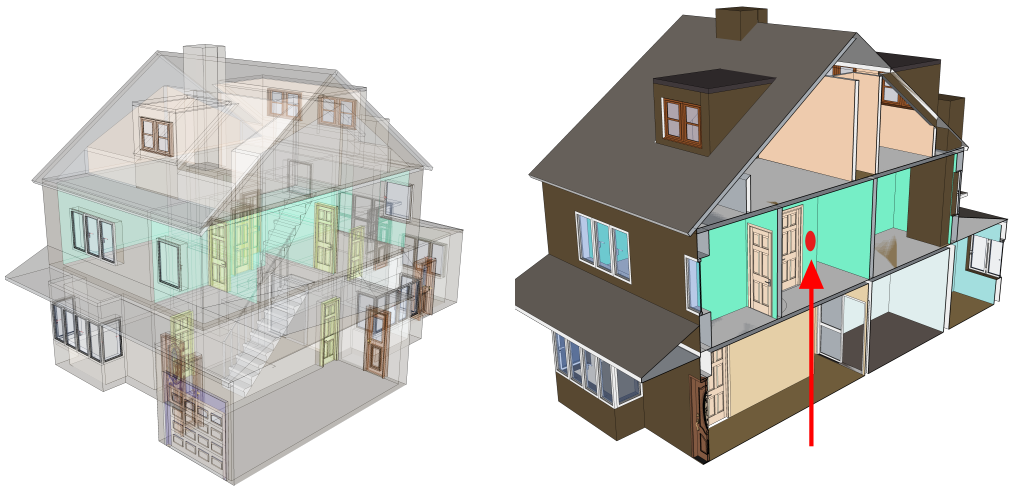
Chapter 6 Space conditioning experiments

In this Chapter the residential home heating and cooling experiment is described. The aim of the experiment was to create prediction models from the recorded data which would model heating and cooling characteristics of a residential home.

The initial approach was to cycle several times through cooling (natural ventilation), and heating of the residential home. Time constraints and structural changes inside the residential home restricted the experiment to one cycle without reaching settled temperatures.

The Chapter begins with a short analysis of the recorded data, in particular the first floor landing air and core (wall) temperatures. It is followed by an iteration through possible curve fitting models and suitable models are presented and discussed.

A 3D model of the house was created by an undergraduate student (Gibbons, 2011) for visualization purposes and a front view from this model can be seen in fig. 6.1. The figure shows the layout of the house and the sensor locations for core and air temperature within the residential home.



(a) *Skeleton view*

(b) *Section view with red dot marking core and landing air sensor location*

Figure 6.1: *3D Model of the residential home*

6.1 Experimental Layout

The experiment started on the 21st October 2011 with forced cooling of the house. The windows and doors were opened to accelerate the temperature drop and to increase the temperature difference

prior to the heating part of the experiment. After inside air temperatures settled and matched the outside air temperatures, all the windows were closed and the heating was turned on continuously for 48 hours and then switched off for 42 hours. In both parts of the experiment, temperature values did not settle at a final value because the experiment was too short. Due to the inhabitants' and the house's security, events that would affect the recorded data were logged. The changes included the opening and closing of doors and windows of the house before the heating part of the experiment, which helped to enlarge the temperature difference from the start of the heating part. Table 6.1 shows a log of the experiment with significant events during the experiment. A complete log can be found in the Appendix in table A.1 on page 142, detailing the actions taken during the experiment.

Table 6.1: *Shortened log of significant events during the heating and cooling experiments conducted in a residential home*

Log ID	Timestamp	Action/Comment
1	2011-10-21T10:00	Boiler turned off - forced cooling
16	2011-10-23T14:15	Boiler turned on - heating
18	2011-10-25T14:15	Boiler turned off - cooling
19	2011-10-27T08:00	End of experiment

6.2 Analysis and considerations regarding experimental data

Experimental data was collected over a period of six days. This is presented in this Section with a discussion of possible improvements for further models. Figure 6.2 highlights the significant events during the experiments taken from table 6.1 alongside both the landing air and core temperature.

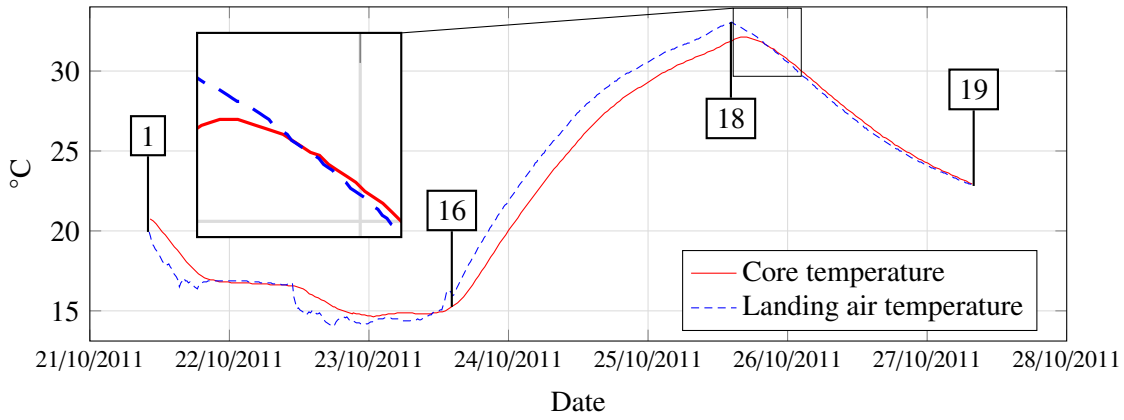


Figure 6.2: *Complete experimental data showing landing air and core temperature with significant events and a zoomed-in segment, showing the begin of the cooling part*

The plot showed some characteristics that were found in the collected data and the following observations were made:

- During the heating part, the core temperature was lower than the landing air temperature.
- After turning off the heating, the core temperature rose for a period of time and the air temperature fell rapidly.

- The rate of cooling for both landing air and core temperature measurements changed during the cooling, with the crossover of core and air temperature being the turning point. See fig. 6.2 at ID '18'.
- In general the core temperature lagged behind the air temperature in response to events.
- The main heating and cooling curvatures for the landing air and core temperature showed a close resemblance to exponential functions and space conditioning experiments carried out under laboratory conditions, see Section 6.4.
- The highest and lowest temperatures during the experiment were significantly different compared to the historical data recorded for the core and landing air temperature sensors, see table 6.2 as a reference.

Table 6.2: *Comparison of historical and experimental extreme temperature values*

Sensor	T_{max}^{Hist}	T_{max}^{Exp}	T_{min}^{Hist}	T_{min}^{Exp}
Landing air	25.8750 °C	33.0625 °C	16.7500 °C	14.0625 °C
Core	25.2500 °C	32.1250 °C	17.0625 °C	14.6250 °C

The following observations were investigated and addressed in this Chapter:

- Outside influences, for example solar radiation, wind, and rain, were neglected. These influences had an impact on the experiment and a correlation analysis of sensory data could find essential relationships between outside phenomena and inside effects.
- The examination and definition of the observed core temperature delay.
- The creation of a model that described both heating and cooling with an exponential model due to the close resemblance of the recorded data.
- The relationship between landing air and core temperature should be investigated.

The investigations into the relationship and related aspects regarding landing air and core temperature are conducted in a single Subsection. The analysis into the outside influences is addressed in a separate Subsection.

6.2.1 Outside influences

It can be seen in fig. 6.2 that landing air and core temperatures showed up and down swings before the heating part, which were induced by the logged events and outside influences. The outside influences were the only source of heat input or cooling, compared to the actual experiments where heating and natural cooling took precedence before external influences.

Outside external influences that could have had an impact during the experiment are summarized in table 6.3 on the following page.

To quantify the effect of the outside influences during the experiments, a correlation analysis would have identified sensors that correlated closely to the observed changes in landing air and

Table 6.3: *Possible outside influence gains and losses during the experiment*

Influence	Heating	Cooling
Solar radiation	✓	-
Rain	-	✓
Wind	-	✓
Air temperature	✓	✓

core temperature. The correlation analysis was not carried out in this Dissertation because of time constraints.

6.2.1.1 Regional climate

The residential home was located within southern England, where the weather can be subjected to continental weather influences (The Met Office, 2014) due to the region being closest to continental Europe. The effects of this were seen during cold spells in winter and hot humid weather during the summer. The temperatures in southern coastal areas were affected by sea breezes, which resulted in lower maximum temperatures and milder temperatures during winter. The winds in this area were not as strong when compared to western and northern Britain, areas closer to the Atlantic. The strongest winds were recorded during winter, between December and February.

6.2.2 Relationship between landing air and core temperature

The landing air and core temperature sensors were placed physically close, therefore a close relationship was anticipated and an analysis of the experimental data was considered to find fitting models.

An essential relationship between landing air and the core temperature is shown as a block diagram in fig. 6.3. The relationships were derived from the initial plot of experimental data in fig. 6.2, from observations, that the landing air affected the core temperature, as convection was the prime source of heat for the core temperature. Conduction of heat through the brickwork from the ground floor was a possible heat input, due to direct heat radiation from adjacent radiators, which were not present at the first floor. The air temperature was consistently higher than the core temperature during the heating. During the cooling, the core temperature lagged behind from the point on where core and air temperatures crossed. The core temperature followed the air temperature during both the heating and the cooling phase.

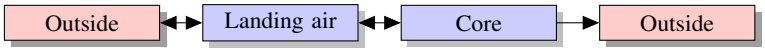


Figure 6.3: *Diagram of the relationship between core, landing air and the outside, showing energy exchanges as arrows during times when the outside air temperature is lower than the inside air temperature*

The observations on the temperature difference between landing air and core temperature were plotted as temperature difference $\Delta T = T_{air} - T_{core}$ and can be seen in fig. 6.4.

The maximum and minimum values of ΔT for the whole historic data were similar to the experimental data. This can be seen in fig. 6.5 for the historic data and the extreme values between 'Heating start' and the end of the experiment can be seen in fig. 6.4. The experimental curve showed

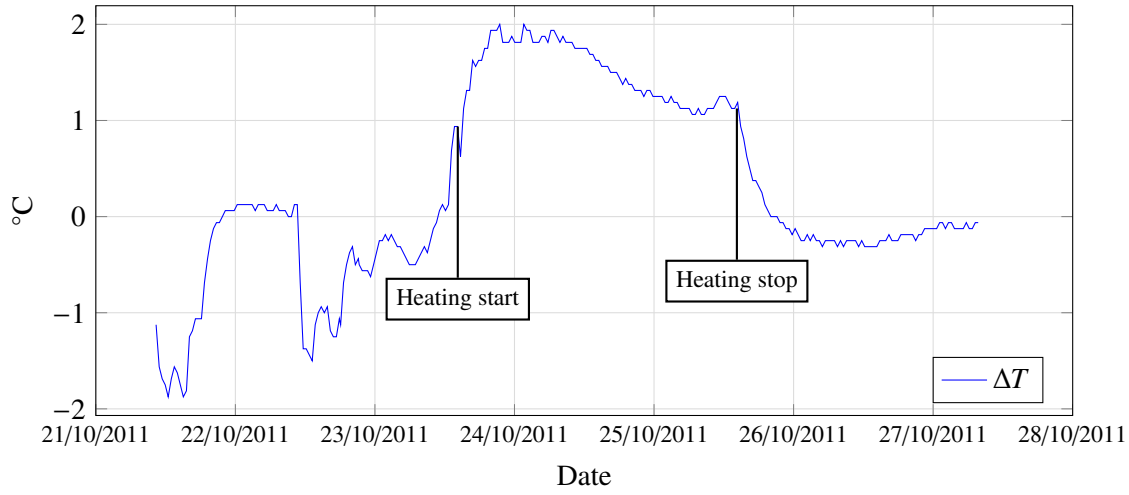


Figure 6.4: *$\Delta T = T_{air} - T_{core}$ plotted for whole experiment*

a sharp incline in ΔT at the 'Heating start', which was due to a higher landing air and lower core temperature. The same effect can be seen in fig. 6.2 on page 88, where the air temperature changed faster than the core temperature.

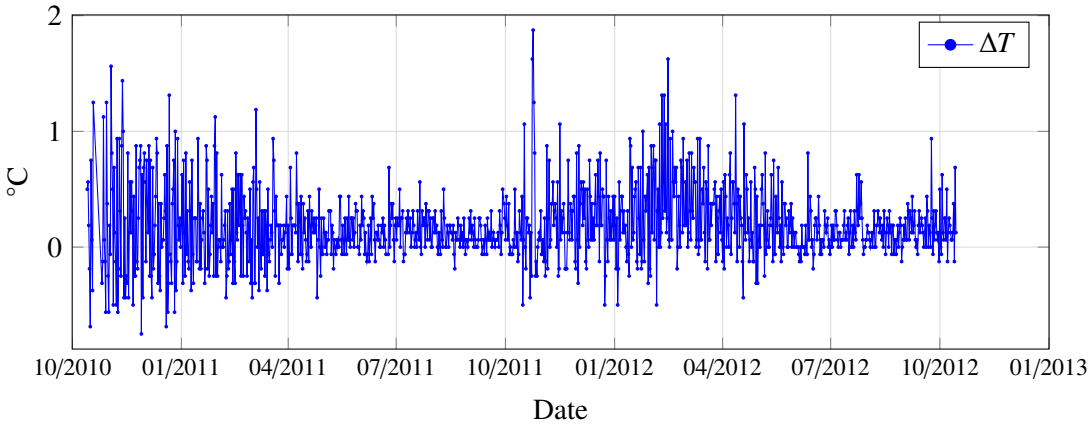


Figure 6.5: *$\Delta T = T_{air} - T_{core}$ plotted for whole historical data, showing similar extreme values compared to the experiment*

The heating and cooling ΔT showed a similar pattern, a sharp incline/decline in the beginning and then a decreased curvature. As the experiment was shorter than 48 hours for each cycle insufficient data was available to analyse ΔT further.

The delay of the core temperature could be described with the materials involved: Changes in temperature, excited by the internal heating, propagated primarily by convection in the air and by conduction inside the brickwork, which was slower. The finding of datasets within the historical data resembling the experimental data would help in a further analysis. Therefore the historical data was analysed with the sensor *HeatedHouse*. The temperature data collected after the heating was turned off helped to observe the cooling by natural ventilation. Datasets with a close resemblance to the experiment were found during night time, where no external heat gains were witnessed. Due to the heating demand stopping at different times, only a few long hours of cooling were observed.

Therefore the relationship between the two sensors was not further investigated.

6.3 Experimental models

The recorded data during the main cooling and heating part of the experiment had a close resemblance to exponential functions. This Section details the initial models created during the research and the improved models developed from them.

6.3.1 Simple exponential models

The initial data for both landing air and core temperature indicated the application of a basic exponential function as a baseline curve fit. The formulas used to model heating and cooling were as follows:

$$f(t)_{heat} = a(1 - e^{-bt}) \tag{6.1}$$

$$f(t)_{cool} = e^{-at} \tag{6.2}$$

Achterbosch et al. (1985) conducted space conditioning experiments on residential homes to verify their inside air prediction models. The inside temperatures in their experiments settled at final temperature values after several days, as can be seen in fig. 2.1 on page 15.

The full heating and cooling cycles were not obtained during the space conditioning experiment on the residential home described in this Chapter. However some datasets were extended by curve fitting to show expected final temperature values.

The core and air temperatures for the heating experiment were subtracted by their respective starting temperatures to make 0 °C the starting temperature. This step was carried out to make curve fitting possible for eq. (6.1) without the use of an offset temperature. The core and air temperatures did not have the same temperature values at the beginning, but were plotted side by side for comparison.

In the cooling experiment, the final temperature value was selected as the outside air temperature recorded at the end of the cooling experiment. The inside temperatures would settle at a value equal to the outside air temperature, if given sufficient time. The cooling experimental data was normalised to a range from ‘1’ to ‘0’, with ‘1’ representing the starting temperature and ‘0’ as the outside air temperature.

From this data, curve fits were created and a least squares error function used to create a fitting model with parameterized values. The graphs in figs. 6.6 and 6.7 show the results obtained with this method and the parameters found.

6.3.1.1 Discussion

Due to time constraints, there was not enough data to fully confirm the initial exponential models. Future experiments should take into account longer settling times for closer final temperature values, for example three to six days.

A higher order response model may have been more suitable at the beginning of the cooling part of the experiment. The higher order was not addressed with this basic initial model. The reason for the slight increase in temperature at the beginning was due to heat emitted from the solid core in

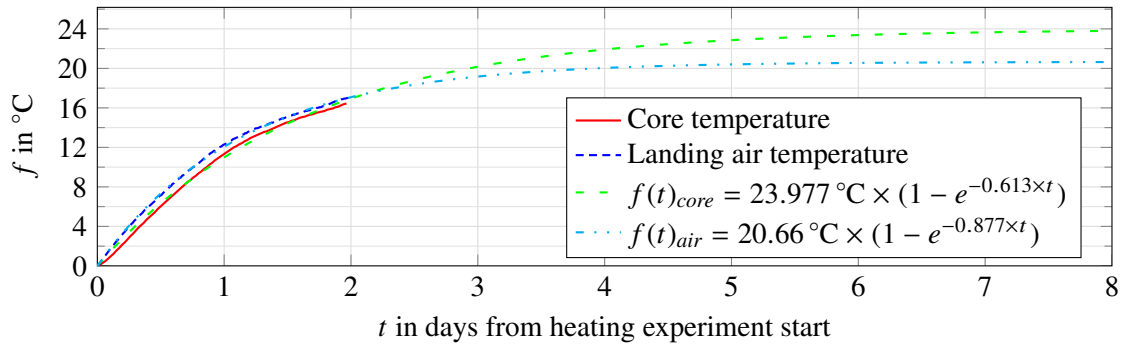


Figure 6.6: *Heating part with start temperatures subtracted from the experimental data and the model data*

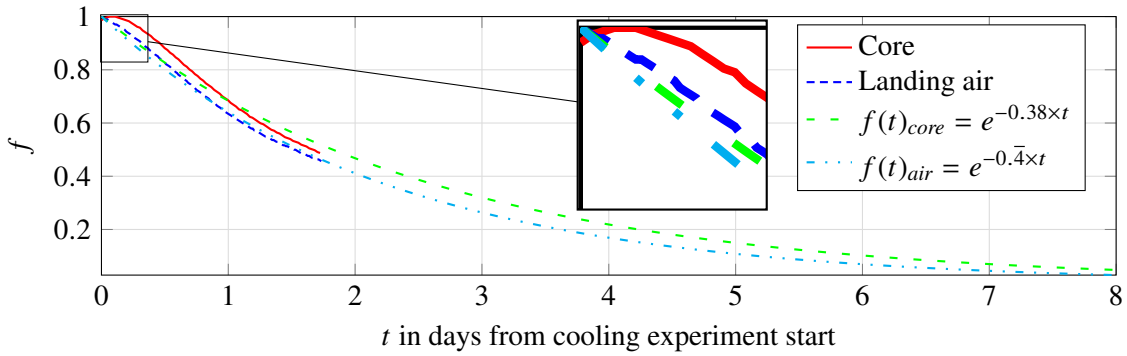


Figure 6.7: *Cooling part with normalised experimental data and model data for the landing air and core temperature, showing a zoomed-in part of the plot at the beginning*

the residential home. This resulted in the air temperature decreasing in general, but damped due to secondary heating introduced by the surrounding brick walls. This can be seen in a zoomed-in segment of the plot in fig. 6.2 on page 88, showing air and core temperature at the begin of the cooling part of the experiment.

The models fitted to the heating experimental data showed a better alignment. Due to the limitations of the basic initial exponential models, new models were proposed and they are described in the next Subsection.

6.3.2 Improved data and different model types

In a second approach, higher order models were used to fit the whole cooling data. The model types selected to fit the complete cooling experiment data were a fourth degree polynomial and a two parameter exponential function. The polynomial function prototype is shown in eq. (6.3) and the exponential function shown in eq. (6.4).

$$f(t) = at^4 + bt^3 + ct^2 + dt + g \tag{6.3}$$

$$f(t) = ae^{bt} + ce^{dt} \tag{6.4}$$

Coefficients were named a, b, c, d, and g. The results for the fits can be seen in tables 6.4 and 6.5, which detail the parameters found and the quality of fit with R^2 and Root Mean Square Error (RMSE) values. The cooling experimental data was normalised to a range from '0' to '1' and

the heating experimental data was subtracted with the starting temperatures, starting from 0 °C, as described in Subsection 6.3.1 on page 92. The graphs of the polynomial and exponential fits can be seen in figs. 6.8 and 6.9.

Table 6.4: *Curve fitting polynomial models for core and air temperature*

Curve fit	a/d^{-4}	b/d^{-3}	c/d^{-2}	d/d^{-1}	g	R^2	RMSE
Heating Air	0.8747 °C	−2.657 °C	−1.755 °C	15.66 °C	0.097 54 °C	0.9998	0.0614 °C
Heating Core	1.47 °C	−6.061 °C	4.78 °C	11.32 °C	−0.1883 °C	0.9999	0.043 39 °C
Cooling Air	-0.7704	0.3363	-0.4078	-0.2156	0.9994	1	0.00115
Cooling Core	-0.1423	0.6292	-0.8782	0.07641	0.9965	0.9998	0.002586

Table 6.5: *Curve fitting two-term exponential models for core and air temperature*

Curve fit	a	b/d^{-1}	c	d/d^{-1}	R^2	RMSE
Heating Air	198.9 °C	-0.3233	−199 °C	-0.4118	0.9996	0.097 22 °C
Heating Core	684.9 °C	-0.3085	−685.6 °C	-0.332	0.9988	0.1777 °C
Cooling Air	-0.06208	-8.327	1.056	-0.4976	0.9992	0.00493
Cooling Core	1.138	-0.5033	-0.1545	-6.567	0.9994	0.004034

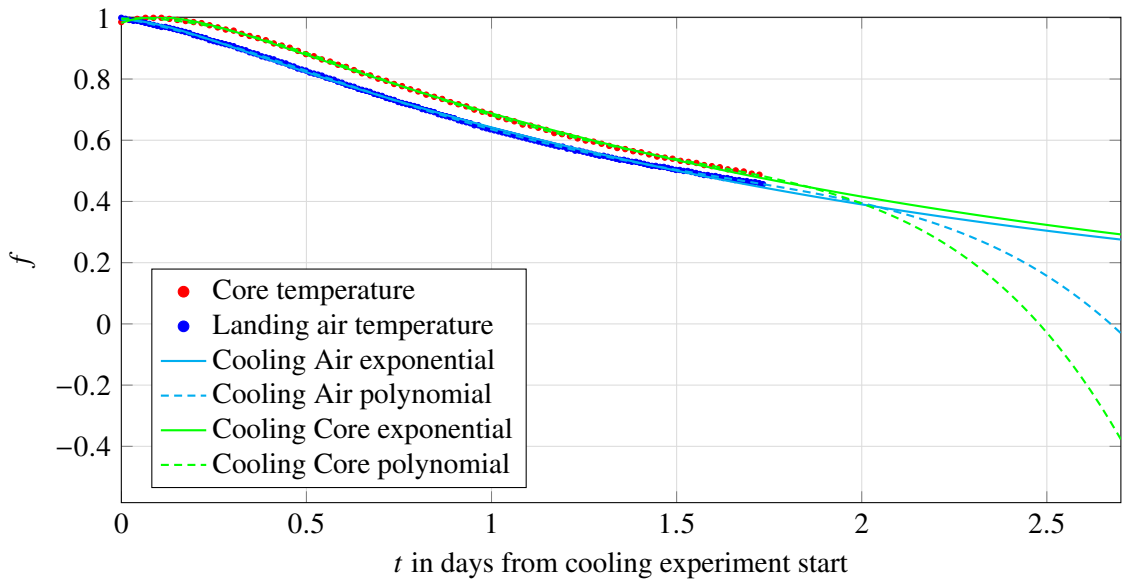


Figure 6.8: *Cooling part of the experiment with normalised experimental data and model data from polynomial and exponential fit*

The results showed a qualitatively good fit of all models. A visual review was carried out to investigate the curve fit beyond the experimental data, and it was found, that the polynomial model only fitted the experimental data. Based on the experimental results by Achterbosch et al. (1985), the polynomial model was found incompatible with the observed exponential curvature. Investigations into the polynomial models were therefore discontinued and only the exponential models considered for further curve fittings.

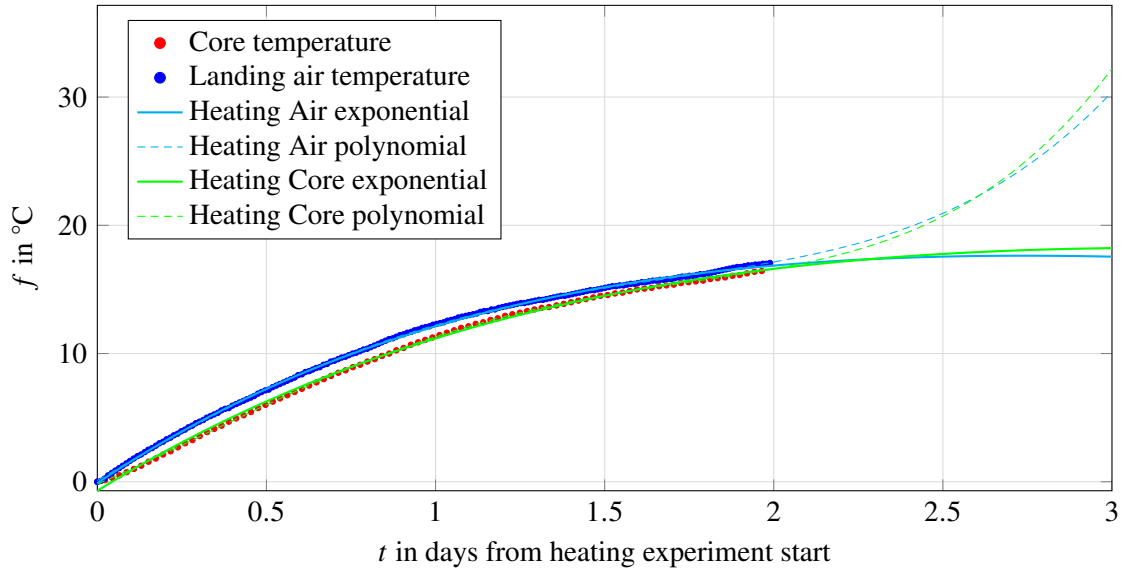


Figure 6.9: *Heating part with start temperatures subtracted from the experimental data and model data from polynomial and exponential fit*

6.3.3 Improved exponential models

The new models with their improvements showed a significantly better fit than the basic exponential model but lacked parameters related to temperatures. Therefore two extended exponential formulas were used for cooling and heating, incorporating temperature based parameters. The formulas are described in eqs. (6.5) and (6.6).

$$f(t)_{cool} = a + be^{ct} \tag{6.5}$$

$$f(t)_{heat} = a + b(1 - e^{ct}) \tag{6.6}$$

The parameters for the cooling model were:

a was the end temperature reached at infinity T_∞.

b was the temperature difference from start to end temperature $T_{start} - T_\infty$.

c was the specific time constant with unit per day (d^{-1}).

t was the time in days.

The following parameters were different in the heating equation:

a was the start temperature T_{start}

b was the temperature difference from infinity to start temperature $T_\infty - T_{start}$

Data used for the cooling model fits was offset from the beginning to find a better fitting model. The time offset was selected due to the effect observed for the core temperature exhibiting a higher order response, which is discussed in Subsubsection 6.3.1.1 on page 92. Three models each were

Table 6.6: *Curve fitting models for cooling of core and air temperature*

Name	Offset	a/°C	b/°C	c/d^{-1}	R^2	RMSE/°C
Cool1Air	-	11.31	22.26	-0.3926	0.9974	0.1646
Cool1Core	-	-24.6	57.59	-0.1156	0.9915	0.2834
Cool2Air	6 h 15 min	16.93	17.6	-0.6381	0.9994	0.0648
Cool2Core	6 h 15 min	14.41	19.93	-0.5006	0.9989	0.0866
Cool3Air	10 h 45 min	18.11	17.03	-0.7424	0.9997	0.0387
Cool3Core	10 h 45 min	16.86	18.28	-0.6415	0.9996	0.0469

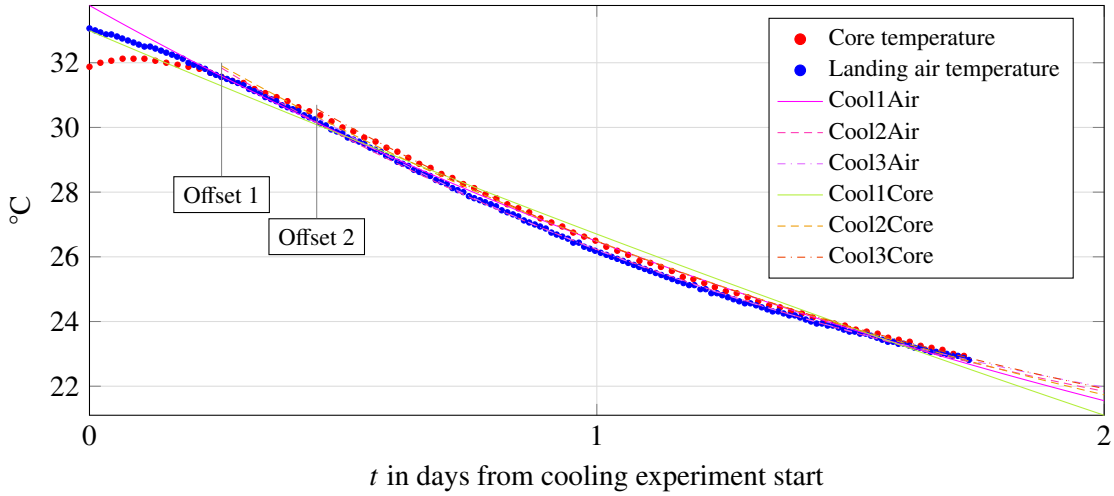


Figure 6.10: *Cooling part of the experiment with experimental data and fits*

created for core and air temperature. The parameters and model comparison criteria can be seen in table 6.6. Plots of the experimental data and the models can be seen in figs. 6.10 and 6.11.

The models' quality measures did not change significantly compared to the parameters. The *Cool1Air* and *Cool1Core* fits did not yield reasonable temperature values as parameters and were therefore discarded. The *Cool2Air*, *Cool2Core*, *Cool3Air* and *Cool3Core* curve fits showed consistent temperature parameters and had increased R^2 and decreased RMSE values.

Newton's law of cooling could also be considered with the results. It stated that, given the

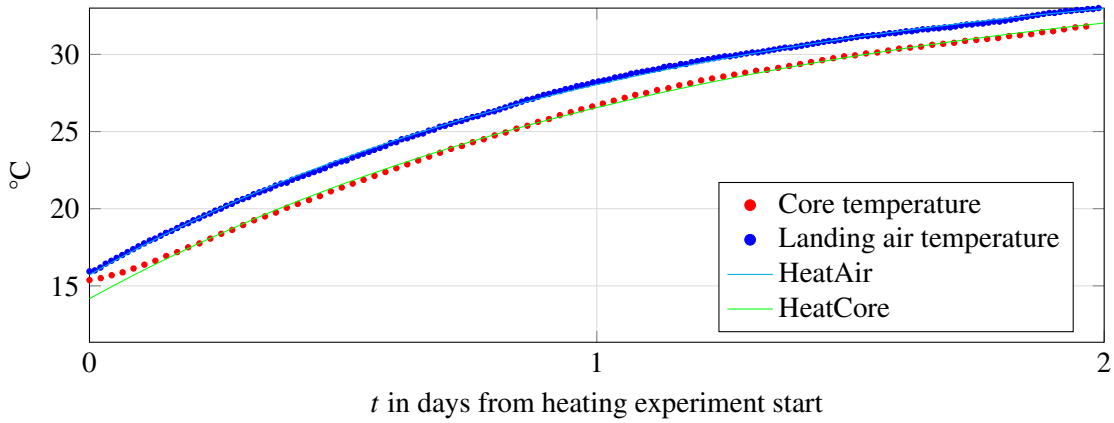


Figure 6.11: *Heating part of the experiment with experimental data and fits*

conditions were the same and the cold and hot environment were exchanged, the same parameters should be yielded for the heating and cooling exponential functions (Desai, 2006). In the residential home, the environment consisted of two materials in contact: a solid (the brick wall) and a fluid (the air). During heating and cooling, the two experiments should yield the same characteristics in terms of their curve fitted parameters. With these conditions, the exponential heating model in eq. (6.6) was used and the results can be seen in table 6.7 for both landing air and core temperature.

Table 6.7: *Curve fitting models for heating of core and air temperature*

Name	a/°C	b/°C	c/d^{-1}	R^2	RMSE/°C
HeatAir	15.71	20.55	-0.9164	0.9994	0.1153
HeatCore	14.17	22.22	-0.8146	0.9988	0.1752

6.3.3.1 Discussion

The results were not conclusive when considered against Newton's law of cooling. The causes for this could be found within the execution of the experiment, where outside influences were neglected. The outside influences could be controlled in an experiment conducted indoors or in a laboratory setting. As there was only one cycle each for heating and cooling, this only left one set of each to compare with. A series of experiments could have led to a more accurate average and comparison of cooling and heating temperature observations.

6.3.4 Effect of parameter *c* on model fit results

The results shown in tables 6.6 and 6.7 on page 96 and on the current page showed a conclusive result in parameter c: the core temperature changes were slower than the landing air temperature. To quantify the changes of temperature, the heating and cooling exponential models were simplified to the main exponential formulas:

$$f(t)_{cool} = e^{ct} \tag{6.7}$$

$$f(t)_{heat} = (1 - e^{ct}) \tag{6.8}$$

In eqs. (6.7) and (6.8), $f(t)$ did not represent a temperature, but the percentage of the temperature difference between the start and the end temperatures. For cooling, this represented the remaining percentage and for heating, the increasing percentage of the overall temperature difference. The c parameters found during the curve fitting of *Cool3Air*, *Cool3Core*, *HeatAir* and *HeatCore* were compared in fig. 6.12.

In this plot, the intermediate percentages calculated with eqs. (6.7) and (6.8) are shown in highlighted boxes for one, two, three, four, five and six hours for each model's c parameter. The percentages represented increases for the heating and decreases for the cooling. 100% represented the total temperature difference between $t = 0$ and $t = +\infty$

As can be seen in fig. 6.12 the change in temperature was lower than 20% of the total temperature difference during the first six hours. As this percentage gave an indication of where the relative

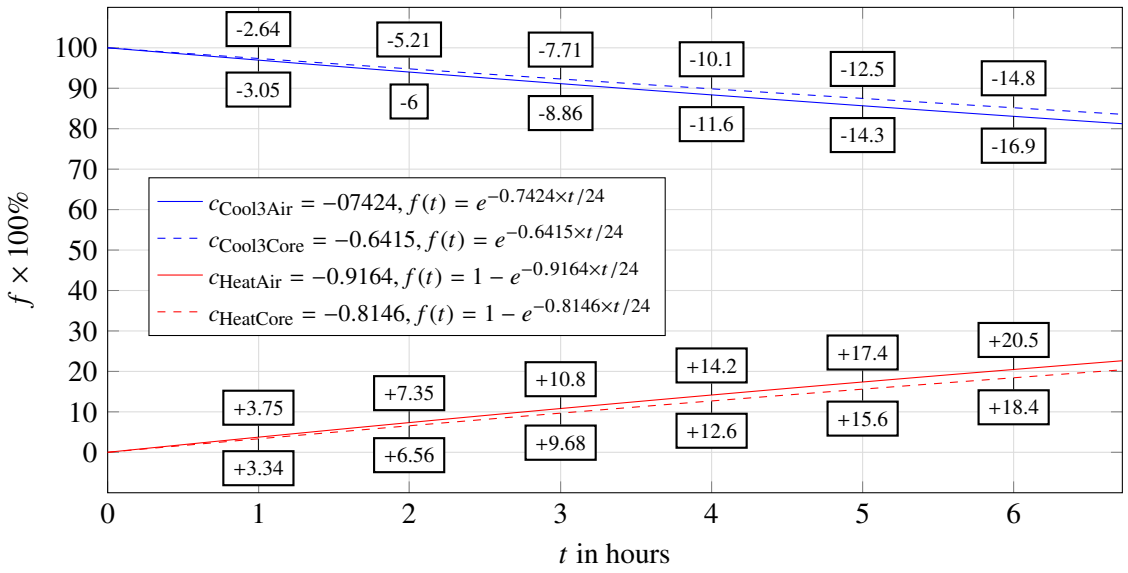


Figure 6.12: *Comparison of the c parameters from different curve fits and display of the increase and decrease of total temperature difference in percent*

end temperatures would be, this was used to further analyse core and air temperature in the historical data to find missing parameters in eqs. (6.5) and (6.6). Using the formulas eqs. (6.7) and (6.8) and the maximum hours of a dataset revealed the increase or decrease in the temperature difference. For example if the landing air temperature was 21 °C at the beginning of a heating period and 2 hours later at 21.5 °C, then this increase was 7.35% of the total temperature difference of 6.8 °C.The same method was applied to vacancy datasets found within the Golden Dataset described in Subsection 4.5.4, taking only heating and cooling datasets. Then the total difference as the percentage was calculated, which was known by the length of the dataset and the final temperature value. The temperature difference between start and end of a dataset was used and multiplied to calculate the total temperature difference. With this approach following results were acquired from these sets:

- Maximum cooling temperature difference: −4.1571 °C (Landing air) and −3.1764 °C (core).
- Minimum cooling temperature difference: −0.3369 °C (Landing air) and −0.3301 °C (core).
- Minimum heating temperature difference: 0.3402 °C (Landing air) and 0.3701 °C (core).
- Maximum heating temperature difference: 3.4504 °C (Landing air) and 2.7610 °C (core).

6.3.5 Stretched exponential models

The improved exponential models did not fit the delays shown by the core temperature. Therefore, other non-linear models were considered and stretched exponentials selected, also called 'Kohlrausch functions'. The definition of stretched exponential models was first stated by Kohlrausch (1854) and the models were applied to experimental data of capacitors consisting of glass jars with metal foil. The transient responses of voltage and current recorded from experiments were then described with stretched exponentials. Recent research has applied stretched exponentials in

different areas, such as biology and chemistry (Anderssen, Husain & Loy, 2004). The 'Kohlrausch function' was used in this Dissertation to improve the exponential model fitting. The 'Kohlrausch function' introduced a power β to the exponential factor, which can be seen in the following equations:

$$f(t) = e^{-t^{\beta}} \tag{6.9}$$

$$f(t) = 1 - e^{-t^{\beta}} \tag{6.10}$$

In general, values of $0 < \beta < 1$ were referred to as 'stretched' and values of $\beta > 1$ were referred to as 'compressed' exponentials, with the time t defined as $0 < t < +\infty$. The names were given to these exponential functions because of their deformation in comparison to regular exponential functions. Figure 6.13 shows normal, stretched and compressed exponentials in comparison, applying the formula described in eq. (6.9).

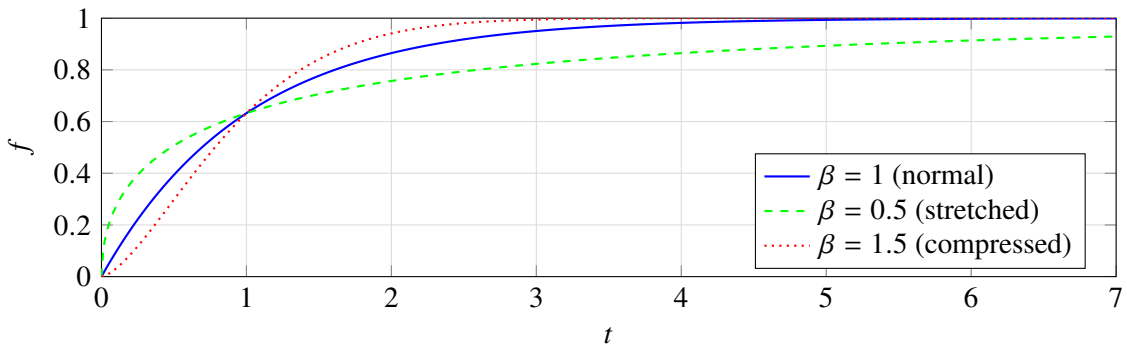


Figure 6.13: *Comparison of normal, stretched and compressed exponential function*

The stretched exponential had a steeper incline and it took longer to reach the maximum. The compressed exponential exhibited a slower incline, but reached the final value earlier.

The exponential models used to fit the experimental data to were defined as follows:

$$f(t) = a + be^{ct^{\beta}} \tag{6.11}$$

$$f(t) = a + b(1 - e^{ct^{\beta}}) \tag{6.12}$$

Both the heating and cooling experimental data were fitted with a single model each, and they retained a qualitatively good fit, compared to the last results described in table 6.6 on page 96.

Table 6.8: *Curve fitting stretched exponential models of core and air temperature*

Name	a/°C	b/°C	c/d^{-1}	β	R^2	RMSE/°C
HeatAirStretch	16.05	19.09	-0.9973	1.090	0.999	0.0837
HeatCoreStretch	15.37	18.1	-0.9825	1.265	0.999	0.0548
CoolAirStretch	20.81	12.16	-0.864	1.418	0.999	0.0316
CoolCoreStretch	21.79	10.43	-0.8846	1.775	0.999	0.0707

The results in table 6.8, show, that both models were a better fit for the experimental data,

showing R^2 values of close to 1 and small RMSE values. Plots of the fits for cooling and heating can be seen in figs. 6.14 and 6.15.

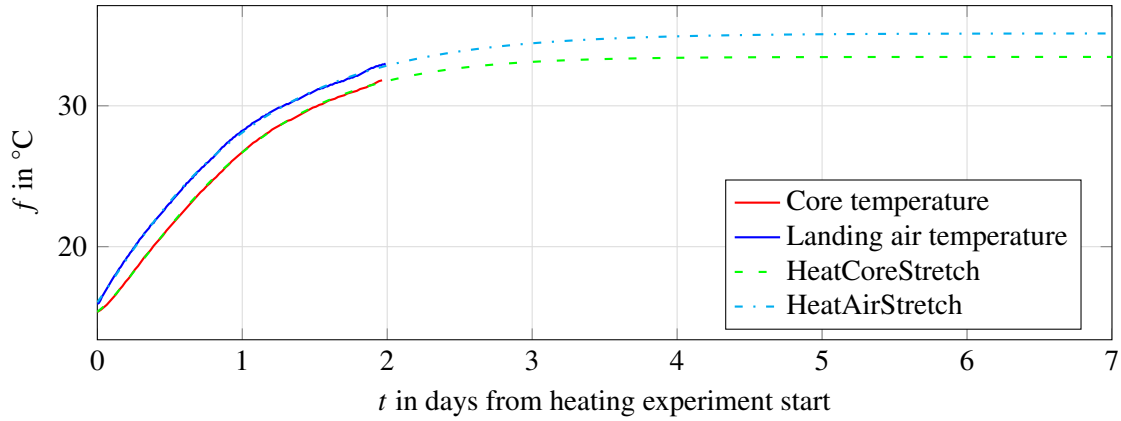


Figure 6.14: *Stretched exponential curve fits and original experimental data for heating part*

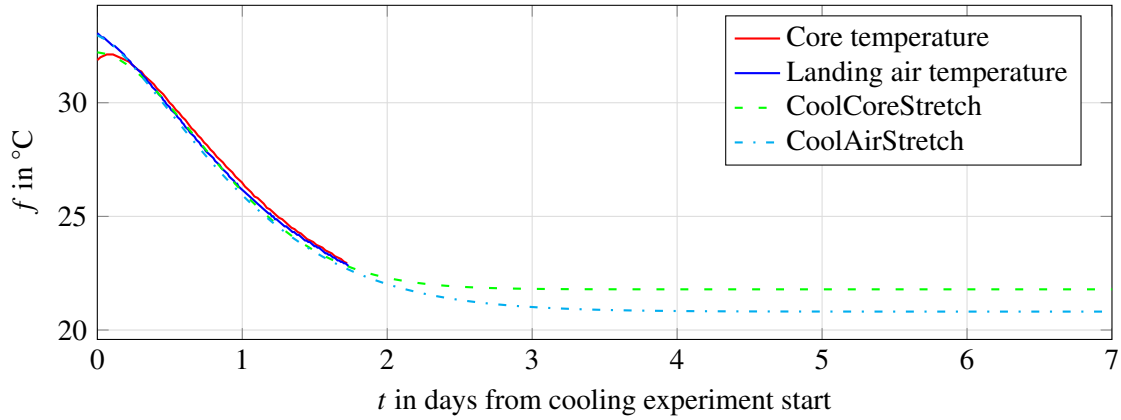


Figure 6.15: *Stretched exponential curve fits and original experimental data for cooling part*

The latest model showed, that the delay by the core temperature could be fitted with compressed exponential models. The different values for β and the parameter c for the heating and cooling models indicated a second time, that Newton's law of cooling was not applicable to the space conditioning experiments on the residential home. Furthermore, the results obtained with ΔT, shown in fig. 6.16, backed up the observation: the temperature difference was not the same for cooling and heating, therefore the energy used for heating was not entirely retained by the core. This can be seen by the larger difference in ΔT for heating than for cooling.

6.3.5.1 Discussion

The stretched models provided the best curve fits for the recorded data during the experiments. They described characteristics resembling the ΔT for the landing air and core temperature, which were discussed in Subsection 6.2.2 on page 90 and plotted in fig. 6.4 on page 91. Plots from heating and cooling stretched exponential models revealed a similar pattern, when compared to the experimental data. The two plots can be seen in fig. 6.16.

The shape of these curves resembled step input responses. However it was not possible to

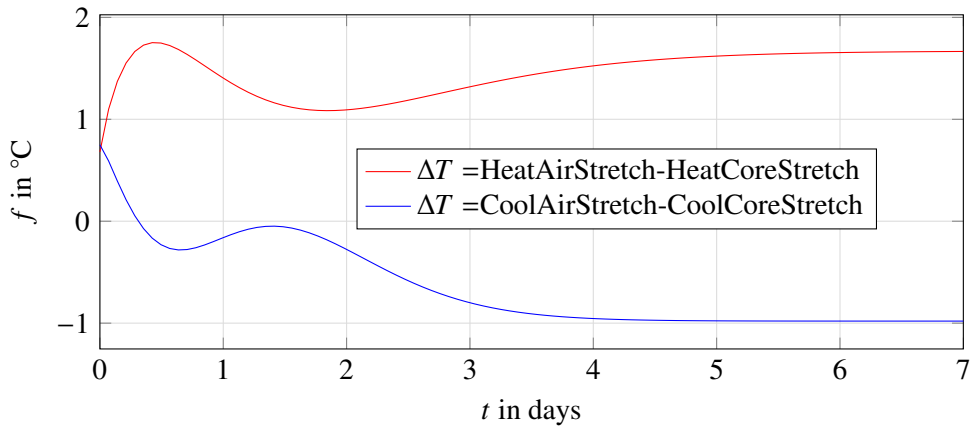


Figure 6.16: *ΔT for stretched exponential models*

fit second order models. In each case, an over-swing and under-swing of the final value was observed, and after six days a stable offset between core and air temperature was reached, with the air temperature being higher.

6.3.6 Core model

The preceding models had separate cooling and heating models. A core model was proposed, which was comprised of different inputs and one output; the core temperature. This model would have used a different approach, in that it would have incorporated a constant cooling part and a heating part. The constant cooling part would have described the heat loss through the envelope of the residential home to the outside, when inside temperatures were higher than the outside temperature. The heating part would have consisted mainly of the heat gains through solar radiation and the internal heating system. The cooling and heating parts would have used the stretched exponentials models, as these described the observed behaviour the best so far, presented in Subsection 6.3.5 on page 98.

The core model design layout can be seen in fig. 6.17, showing inputs, internal components, and output of the core model.

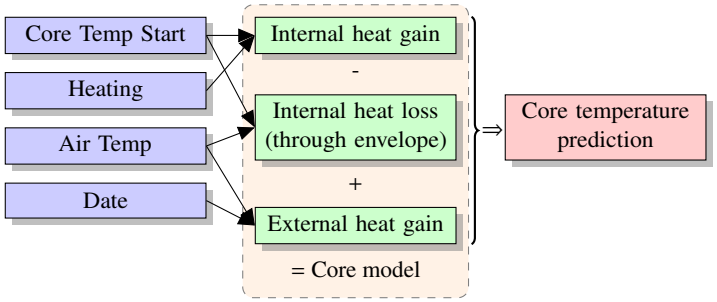
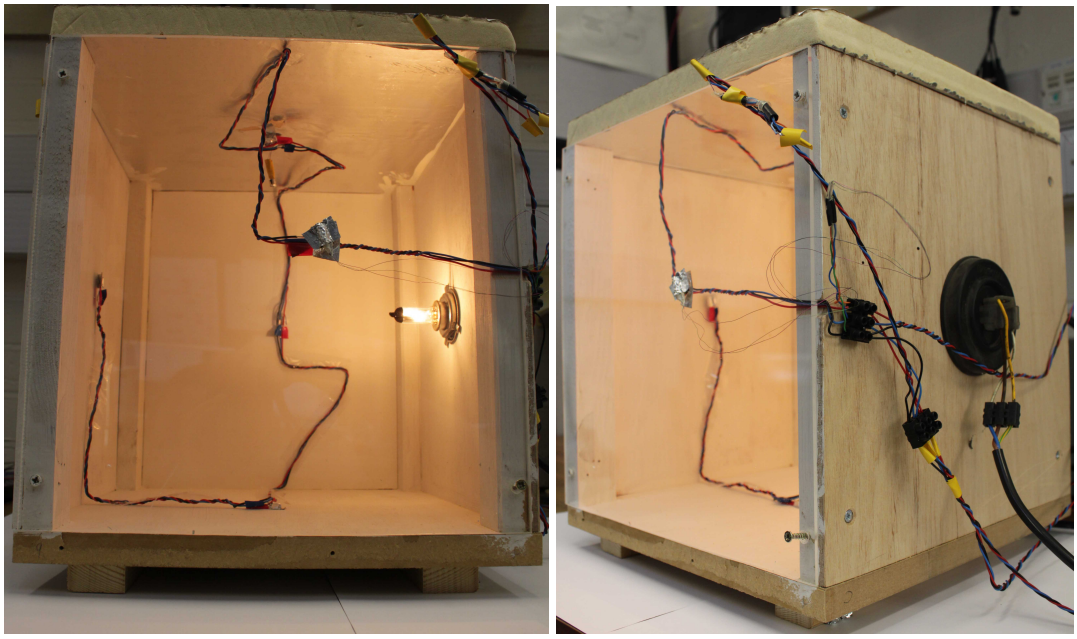


Figure 6.17: *Core model diagram*

Due to time constraints, the design and approach were not tested. If such a connection could be made in future research, an improved model for short term predictions of the core temperature could be found.

6.4 Laboratory space conditioning experiments

During the research described in this Dissertation, undergraduate students (Severne, 2011; Li, 2012) were involved with the development of an enclosed box, which was used for heating and cooling experiments. This box was equipped with a 1-wire network of temperature sensors and a 55 W car bulb as a heat source to simulate the experiments conducted in the residential home. The box was placed inside an environmental chamber and subjected to heating on and off cycles, see fig. 6.18 for pictures of the actual box.



(a) *Front view* ***(b)*** *Side view*

Figure 6.18: *Wooden box with attached sensors and heating source*

The wooden box model was created to simulate the behaviour of heating and cooling and to produce prediction models based on the recorded data to draw parallels to the experiment on the residential home. The laboratory space conditioning experiments on the wooden box were in general carried out as follows:

- The environmental chamber was set to a desired temperature, for example 10 °C, 15 °C and 20 °C.
- The wooden box was left to acquire the same temperature as the ambient temperature.
- The bulb was turned on, initiating the heating part of a cycle.
- The inside air temperature of the wooden box was monitored and the bulb turned off when temperatures settled at a final value.
- The environmental chamber temperature was set to a different temperature and the process repeated.

The same procedure was conducted at 20 °C ambient air temperature in the environmental chamber with one, two, three, and four bricks placed inside the box. One brick included an embedded temperature sensor. A picture of the type of bricks used can be seen in fig. 6.19. The bricks were used to simulate a thermal mass, similar to walls inside a residential home.

Figure 6.19: *Standard Engineering brick with dimensions* $215 \times 102.5 \times 65$ mm

The students' work included an analysis on the collected data applying methods from control theory to create models describing the heating cycle of the experiment. The students' results were not used in this Dissertation and a separate analysis was carried out.

6.4.1 Experimental data

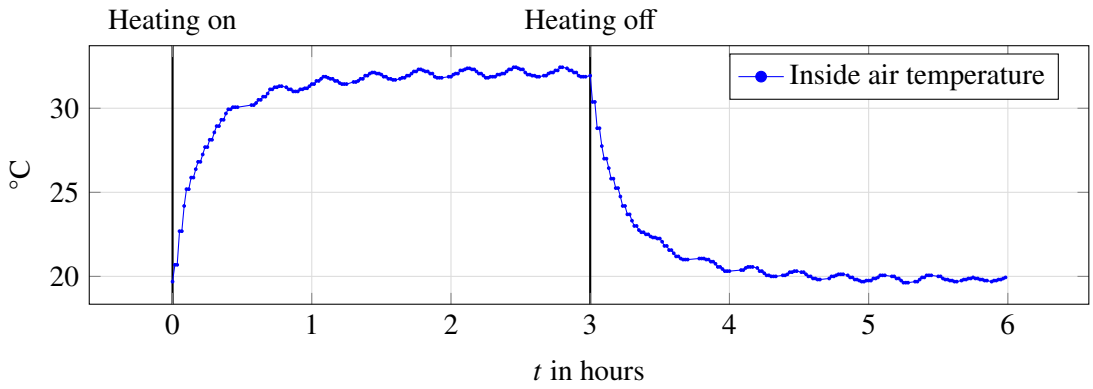
The first set of experiments was conducted in an environmental chamber, which was set to 10 °C and 20 °C, for Cycle 1 and Cycle 2 respectively. The duration for Cycle 1 was 12 hours and consisted of 6 hours each for heating and natural cooling. Cycle 2 was conducted within 6 hours (3 hours for heating and 3 hours for natural cooling).

The second set of experiments in the environmental chamber was undertaken with the introduction of red bricks with a cycle of heating and cooling for each of one, two, three, and four bricks. Plots from the Cycle 2 and with one brick are shown in fig. 6.20. It can be seen in the plot on the top, that temperatures followed a sinusoid progression throughout, which was due to the environmental chamber. The chamber control system applied both heating and cooling to maintain temperatures in reoccurring cycles, which were recorded as sinusoid overlaid patterns in the temperature data. The graph in the bottom does not show the same behaviour. The inclusion of bricks inside the wooden box changed the temperature response and therefore the inside air temperature measurements did not show an overlaid cyclic pattern any longer.

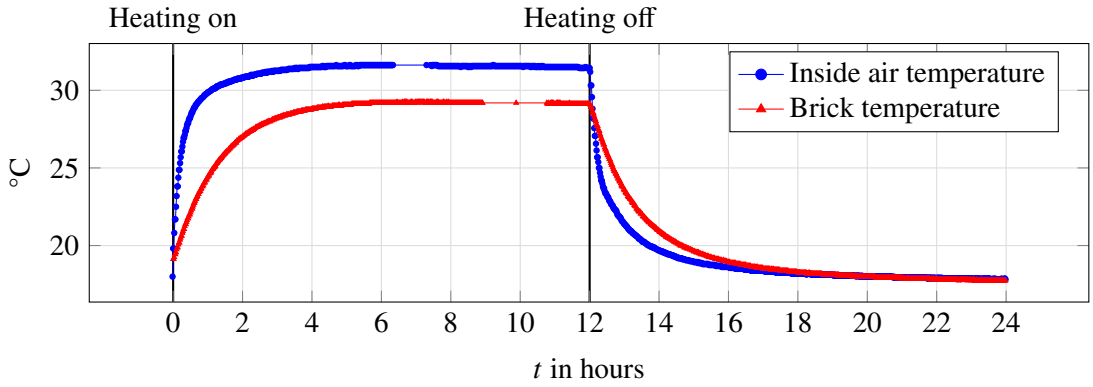
6.4.2 Curve fits

The first curve fits were applied to the inside air temperature of the wooden box. The curve fits for heating and cooling were produced separately. The stretched models, introduced in eqs. (6.11) and (6.12) on page 99, were used to produce these fits. The quality measures R^2 and RMSE were used to confirm the curve fits, with the lowest R^2 value of 0.9872 and highest RMSE of 0.243 °C. The curve fits for cycle one and two can be seen in table 6.9. The curve fits found for experiments including bricks can be seen in table 6.10, the top section for the inside air temperature fits and the bottom section for the brick temperature fits.

Results from the first set of experiments showed a consistent temperature difference between outside and inside air of at least 10 °C irrespective of the environmental chamber temperature



(a) *Inside air temperature*



(b) *Inside air and brick temperatures*

Figure 6.20: *Laboratory space conditioning experimental data recorded at* 20 °C *ambient temperature*

Table 6.9: *Curve fits for inside air temperature for cycle 1 and 2*

Name	T_{ENV}/°C	a/°C	b/°C	c/d^{-1}	β
Cycle 1 Heat	10	8.9	12.7	-34.8	0.7473
Cycle 1 Cool	10	9.4	10.3	-30.5	0.7466
Cycle 2 Heat	20	19.1	13.1	-35.9	0.76
Cycle 2 Cool	20	19.8	11.1	-36.3	0.7896

setting. The parameters for the experimental fits were similar, apart from the start temperature a. The second set of experiments showed different parameter values each time the number of bricks was changed. A significant change was seen in parameter c and β, making the air temperature curve fit stretch further.

The brick temperature curve fits were in general compressed, but showed an increasing temperature difference during heating compared to the air temperature, as can be seen in the bottom plot of fig. 6.20. The temperature differences, taken from the curve fits were 2.3 °C, 2.9 °C, 5.3 °C, and 6.3 °C, for one, two, three, and four bricks respectively.

The effect of bricks on the air temperature can be seen when both brick temperatures and air temperatures were plotted on the same graph. To display and compare the curve fits equally, a was set to '0' and b to '1', leaving c and β unchanged. Four plots of air and brick model comparisons

Table 6.10: *Curve fits for experiments including bricks, with environmental chamber setting of* 20 °C

Name	a/°C	b/°C	c/d^{-1}	β
1 Brick Air Heat	18.5	13.1	-14.6	0.6193
1 Brick Air Cool	17.8	14.6	-7.9	0.5398
2 Bricks Air Heat	17.8	14	-7.6	0.4720
2 Bricks Air Cool	19	13.9	-9.2	0.5685
3 Bricks Air Heat	17.6	15.2	-7.1	0.4779
3 Bricks Air Cool	19	14.2	-6.7	0.5367
4 Bricks Air Heat	18.8	15.6	-6.1	0.5439
4 Bricks Air Cool	18.8	15.6	-5.3	0.5595
1 Brick Brick Heat	19.1	10.1	-20.5	1.043
1 Brick Brick Cool	17.8	11.8	-11.1	0.8584
2 Bricks Brick Heat	18.5	10.5	-11.5	0.881
2 Bricks Brick Cool	19	10.3	-12.9	0.9064
3 Bricks Brick Heat	18.9	8.7	-12.2	1.055
3 Bricks Brick Cool	19.3	8.5	-11.8	1.032
4 Bricks Brick Heat	19.2	8.9	-10.7	1.066
4 Bricks Brick Cool	19.4	8.9	-9.9	1.022

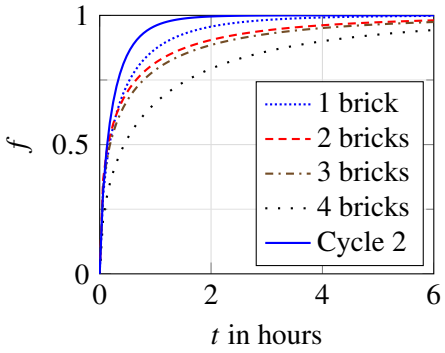
can be seen in fig. 6.21.

The plot shows heating on the left side and cooling on the right side. The top part shows the air temperature models changing with the introduction of additional bricks, including the cycle 2 from the experiments and the air temperature models. Cycle 2 was included because it was conducted under the same environmental chamber temperature setting as the brick experiments. The bottom part shows the brick temperature models. In general, each additional brick increased the time it took to reach a final temperature value for both heating and cooling.

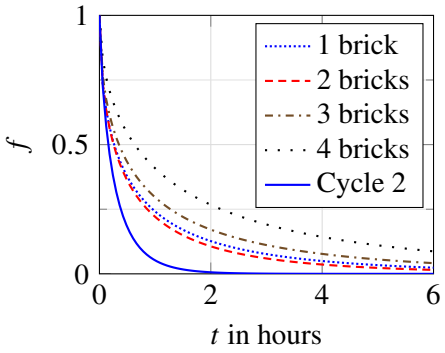
6.4.3 Discussion

The laboratory space conditioning experiments helped to investigate a few aspects about the residential home, which were not possible so far, such as an increase in thermal mass through the introduction of additional bricks and also reducing cycle times from days to hours. The introduction of bricks changed both the brick and the air temperature development. Introducing bricks also increased the settling time for final temperature values. Furthermore the more bricks that were introduced, the higher the final air temperature was, compared to the brick temperature, which remained the same for each configuration of bricks. The first set of experiments revealed that the final inside air temperature had a constant temperature difference to the outside of around 10 °C.

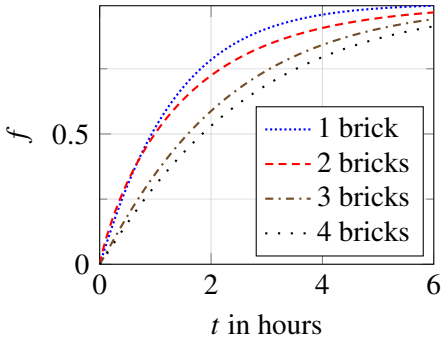
Similarities between the laboratory space conditioning and the residential home experiments can be seen in how brick and air temperatures tended towards different final values, as seen in fig. 6.20. The brick temperature was lower than the air temperature during the heating phase and reversed during the cooling phase. Dissimilarities were seen in the curve fits: the residential home had stretched exponential fits for both air and core temperature and the laboratory space conditioning experiments had compressed exponential fits for the air temperature and stretched exponential fits for the brick temperatures. The differences could be attributed to the simplification of the model that was used to simulate residential home space conditioning.



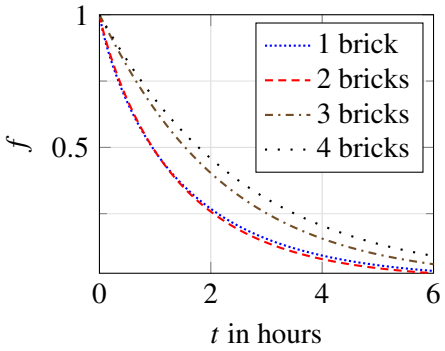
(a) Inside air temperature models for heating



(b) Inside air temperature models for cooling



(c) Brick temperature models for heating



(d) Brick temperature models for cooling

Figure 6.21: *Comparison of normalised laboratory space conditioning experimental model outputs at* 20 °C *ambient temperature, showing an increase in settling time for each additional brick*

6.4.4 Conclusion

The laboratory space conditioning experiments made it possible to take a closer look at inside temperature developments at different outside air temperatures. The important knowledge gained from the laboratory space conditioning experiment observations was as follows:

- End temperatures of brick and air temperatures were different when heating was applied. This was the same behaviour observed during the heating experiment conducted in the residential home.
- End temperatures of brick and air temperatures were the same for cooling. The residential home space cooling period was too short to observe a similar behaviour.
- A constant temperature difference was experienced during heating, irrespective of the ambient temperature.

6.5 Turning curve fits into a model

All models in this Chapter were created by curve fitting to experimental data. The curve fits were parametrized, so that they could be used as general prediction models. The selected parameters based on observations on the experimental, historical and laboratory space conditioning experimental data were as follows:

- The end temperature when internal heating took place.
- The end temperature when cooling took place.

The deductions carried out and the values found for the end temperatures are described in this Section.

6.5.1 Observations on heating end temperature

The end temperature calculated with the stretched exponential model for heating was 32.125 °C and the overall temperature difference between heating begin and end was 16.4375 °C. The core temperature was 15.375 °C when the heating experiment started. Furthermore the following observations were made, affecting the temperature parameters:

- The heating input was consistently the same, the central heating supplied water at 80 °C.
- The outgoing heat loss through the envelope of the house was a function related to the temperature differences of the inside and outside.

If there had been no losses to the outside, in infinite time the inside materials would reach final temperatures, close to the water temperature. But due to losses to the outside, there were limitations to the effects of heating on the residential home. The plot in fig. 6.22 shows temperature differences of the core and outside air temperature at the end of a heating demand, taken from the historical data.

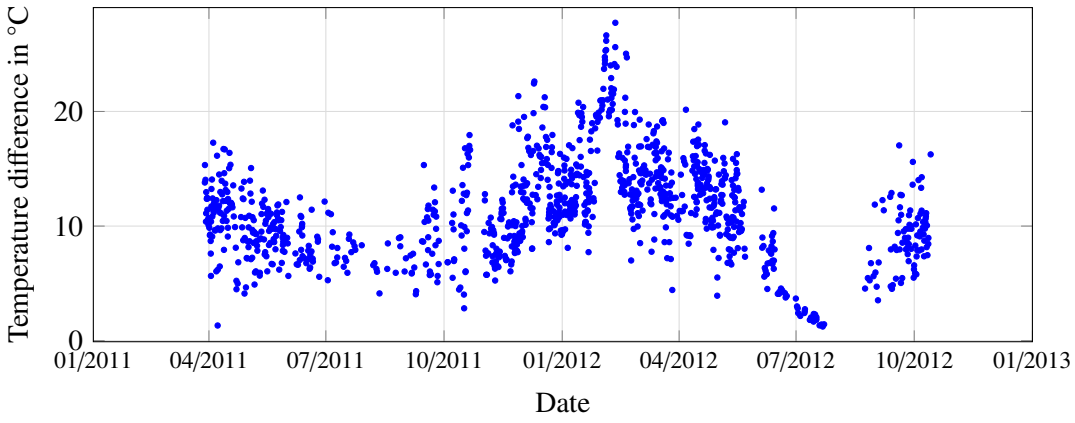


Figure 6.22: *Temperature difference between core and outside air at the end of heating demand sets*

The plot also shows a seasonal change, which coincided with the outside air temperature seasonal change. This plot was used to establish if the experimental models produced a similar offset for the final temperature compared to day to day data and the laboratory space conditioning experiments. The final temperature difference between begin and end was 16.4375 °C, but this did not account for outside temperature changes that occurred during the experiment. A decreased temperature of 12.2 °C was recorded for the outside air temperature at the end of the heating part, resulting in a difference of −3.175 °C for the outside air temperature. The stretched exponential model's predicted core temperature in infinite time was 33.47 °C, the overall difference to the outside was 21.47 °C, if all outside influences would have stayed the same.

The results from the laboratory space conditioning experiment were used to deduce the temperature difference between a final core temperature value and the outside air temperature. A final temperature difference of 21.47 °C was then incorporated into the model and completed the core temperature heating prediction model.

6.5.2 Observations on cooling end temperature

Deductions about the end temperature for cooling of the core were taken from states of equilibrium between the inside air and core temperature. A state of equilibrium was reached, when both inside air and core temperature did not change for at least one hour. A plot showing internal and external temperature difference can be seen in fig. 6.23.

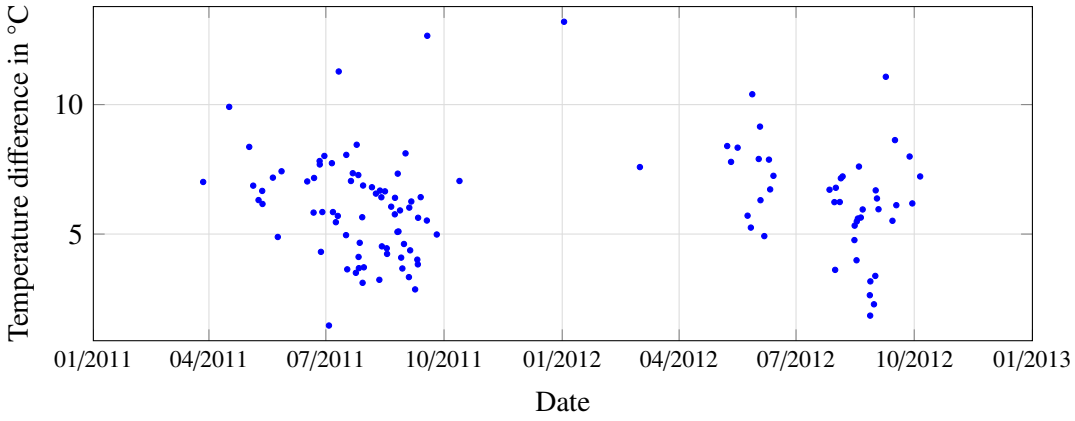


Figure 6.23: *Temperature difference between core temperature and outside air temperature during times of equilibrium between core and inside air temperature*

The equilibrium states were predominantly observed during the summers of 2011 and 2012, when no heating was required. This led to the deduction, that the residential home's envelope was seen as a buffer, which enabled the inside to maintain an offset temperature from the outside for short periods of time. Eventually, temperatures would level (in infinite time). When the envelope of a space was small, heat dissipated quicker, as in the case of the wooden box, or if the envelope was perforated, as in the forced cooling of the residential home through opening all windows. The observed data showed that inside and outside air temperatures equalised faster.

The equilibrium states were between 4 °C and 9 °C, with an average value of 7 °C for all equilibrium sets. The average offset from the outside air temperature was used as a parameter to express the difference of the final core temperature for short term predictions.

The two parameters defined in this Subsection were the temperature difference of 21.47 °C for the core temperature between beginning and ending of the heating and the temperature difference of 7 °C for the core temperature between beginning and ending of the cooling. The parameters were used to create models usable beyond the experiments in the next Subsection.

6.5.3 Stretched exponential models for core temperatures

The final two models incorporated the two temperature values and finalised the step of creating working models that could be used to predict temperatures beyond the scope of the experiment:

$$f(t)_{cool} = (T_{out} + 7\,°\text{C}) + (T_{in} - (T_{out} + 7\,°\text{C})) \times e^{-0.8846\,\text{d}^{-1} \times t^{1.775}} \tag{6.13}$$

$$f(t)_{heat} = T_{in} + (21.47\,°\text{C} - (T_{in} - T_{out})) \times (1 - e^{-0.9825\,\text{d}^{-1} \times t^{1.265}}) \tag{6.14}$$

The parameters were as follows:

T_{out} outside ambient air temperature at the start in °C.

T_{in} core temperature at the start of cooling/heating in °C.

t the time in days, starting from zero when heating/cooling started.

The first exponential parameter in each model was the specific time constant with unit 'per day' (d^{-1}), because the time t was given in days. The second exponential parameter in each model was unit-less.

6.5.4 Creating a model for the landing air temperature

The model creation process for the landing air temperature was similar to that of the core model by finding two fixed parameters to create an approximate model for cooling and heating.

The landing air temperature would have reached a final temperature of 35.14 °C, if continuous heating was applied. The final temperature values was obtained from the stretched exponential curve fit and can be seen in table 6.8 on page 99. The same method for the core model was applied to the air temperature model: the outside air temperature at the end of the heating was 12 °C, the temperature difference for the heating air model was therefore defined as 23.15 °C.

The temperature difference for the cooling air model was slightly smaller compared to the core model, as can be seen in fig. 6.15 on page 100. The value for the end temperature difference for the cooling air model was therefore defined as 6 °C.

The following two air temperature models were therefore formulated:

$$f(t)_{CoolAir} = (T_{out} + 6\,°\text{C}) + (T_{in} - (T_{out} + 6\,°\text{C})) \times e^{-0.864\,\text{d}^{-1} \times t^{1.418}} \tag{6.15}$$

$$f(t)_{HeatAir} = T_{in} + (23.15\,°\text{C} - (T_{in} - T_{out})) \times (1 - e^{-0.9973\,\text{d}^{-1} \times t^{1.090}}) \tag{6.16}$$

Figures 6.24 and 6.25 show both models applied to exemplary data for a period of up to six days.

A stretched exponential model that could be benchmarked against alternative models is shown in fig. 6.26. In this setup the inputs into the model were time, the current inside air temperature and the current outside air temperature. As discussed in the core model creation process, this was a general observation, as outside influences (such as cloud coverage, seasonal changes, wind, and rainfall) could have changed, which in turn would influence outside air temperature.

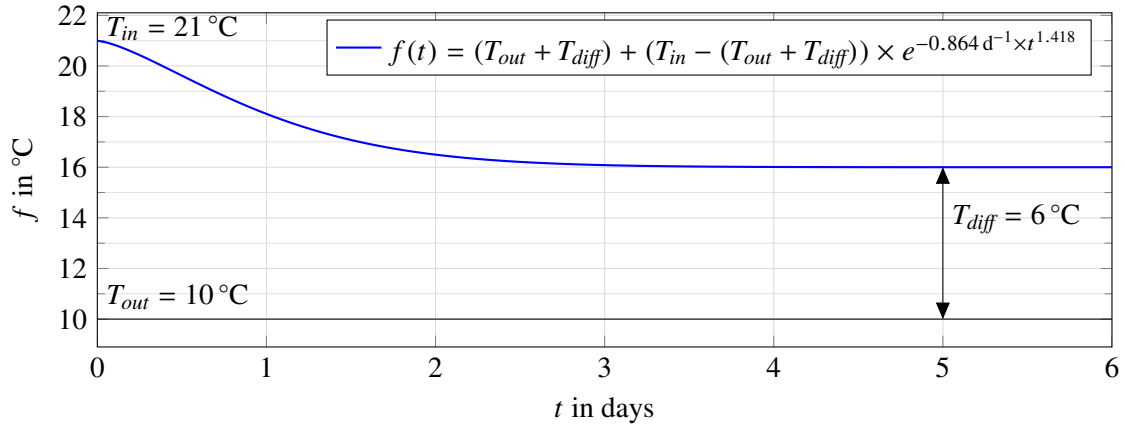


Figure 6.24: *Landing air cooling model applied to exemplary data with* T_{diff} = 6 °C

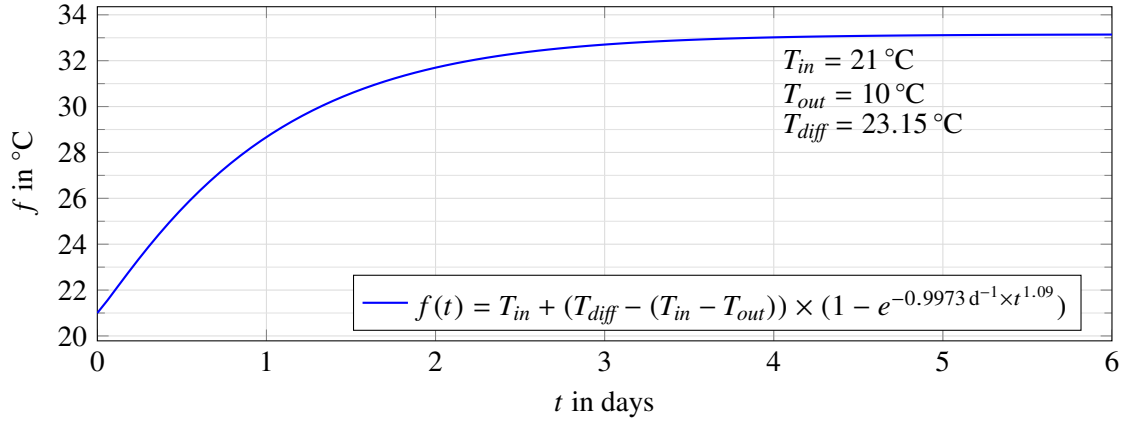


Figure 6.25: *Landing air heating model applied to exemplary data with* T_{diff} = 23.15 °C

6.6 Curve fit and creation of an offset heating model

The heating experimental models were created with results from the laboratory space conditioning experiments. During the laboratory space conditioning experiments both inside air and outside air were at the same temperature before the heating started, which was different to the residential home. The outside air temperature was in general lower than the inside air temperature and in the experimental model approach a subtraction of the air temperatures was used to adjust the exponential model. A different heating model approach was therefore proposed, which started with the same temperatures for inside and outside air and adjusted with a time t_{offset} to align on the x axis

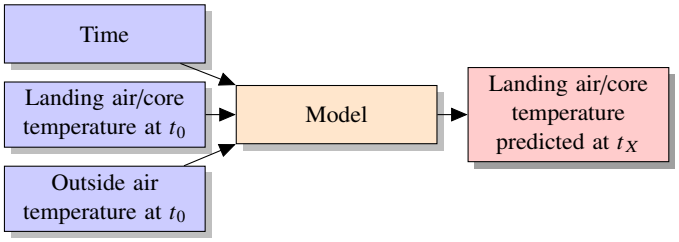


Figure 6.26: *Stretched exponential temperature prediction model*

an intercept for the current inside air temperature as an intermediate point on the exponential model curve fit. The general stretched exponential formula for this offset model approach was defined as follows:

$$f(t)_{offset} = T_{out} + T_{diff}\left(1 - e^{\left(c\left(t+t_{offset}\right)^{\beta}\right)}\right) \tag{6.17}$$

When rearranged at time $t_0 = 0$, t_{offset} was defined as follows:

$$t_{offset} = \sqrt[\beta]{\frac{log\left(1 - \frac{T_{in} - T_{out}}{T_{diff}}\right)}{c}} \tag{6.18}$$

Equation (6.17) was then curve fitted to the experimental data for heating of the residential home, and the values shown in table 6.11 were obtained. The quality measures indicated a qualitatively

Table 6.11: *Curve fitting parameter results and quality measures for the offset model*

T_{out}	T_{diff}	c	β	T_{in}	R^2	RMSE
11.47 °C	23.31 °C	$-0.9139\,\mathrm{d}^{-1}$	1.2	15.97 °C	0.999	0.0837 °C

good fit for the experimental heating air curve fit (as in table 6.8 on page 99). The plot in fig. 6.27 shows the new offset curve fit, old heating curve fit and the experimental data in comparison. The curve fits are as follows:

$$f(t)_{HeatAir} = 16.05\,^{\circ}\mathrm{C} + 19.09\,^{\circ}\mathrm{C} \times \left(1 - e^{\left(-0.9973\,\mathrm{d}^{-1} \times t^{1.09}\right)}\right) \tag{6.19}$$

$$f(t)_{offset} = 11.47\,^{\circ}\mathrm{C} + 23.31\,^{\circ}\mathrm{C} \times \left(1 - e^{\left(-0.9139\,\mathrm{d}^{-1} \times \left(t+t_{offset}\right)\right)^{1.2}}\right) \tag{6.20}$$

$$t_{offset} = \sqrt[1.2]{\frac{log\left(\frac{15.97\,^{\circ}\mathrm{C} - 11.47\,^{\circ}\mathrm{C}}{23.31\,^{\circ}\mathrm{C}}\right)}{-0.9139\,\mathrm{d}^{-1}}} \tag{6.21}$$

The magnified segment of the plot shows the start of the heating experiment, which shows a slight delay of the original landing air temperature data. The *HeatAir* model showed a similar behaviour, whereas the offset model had the same gradient throughout the start. The plots of the offset model compared to the *HeatAir* model can be seen in figs. 6.28 and 6.29.

The first plot shows how the offset model was created and also the heating model in comparison. The second plot shows a zoomed in version of the first plot. The temperature difference between the models did not exceed 0.5 °C in the example data, if the output from the old heating model was subtracted from the offset model. Both models should be compared to actual data when used in prediction, to carry out an analysis assessing if the new offset model was better than the old heating model.

6.6.1 Discussion

The new offset model for heating was created with observations and results from the laboratory space conditioning experiments, where a constant temperature difference T_{diff} was observed between

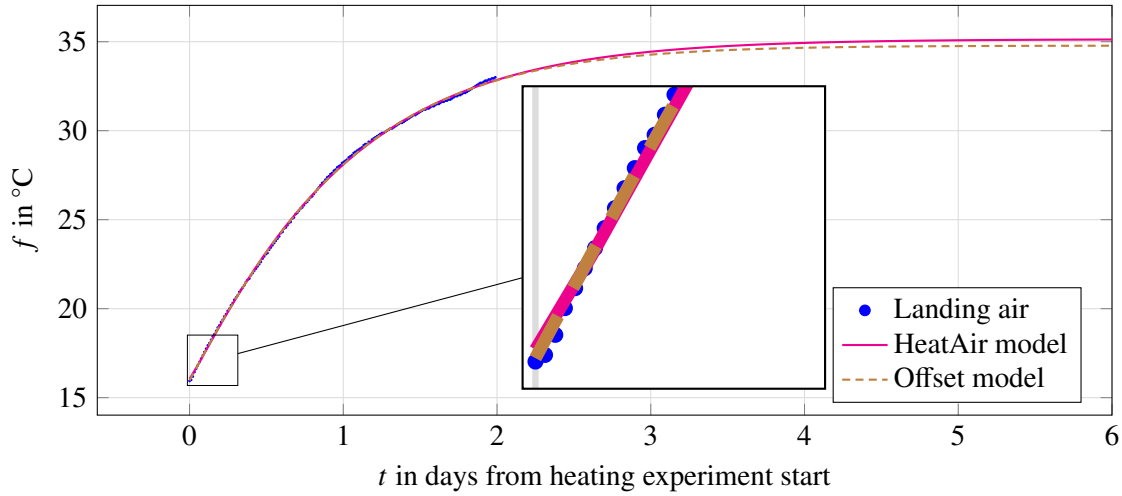


Figure 6.27: *Curve fits for the experimental heating data in comparison*

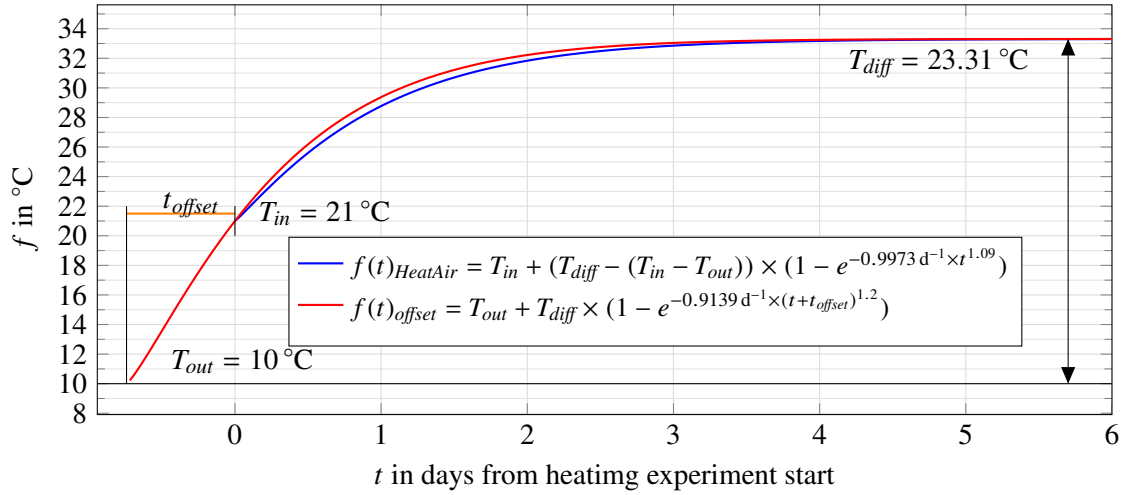


Figure 6.28: *Landing air heating models applied with example values, showing the experimental (red) and offset (blue) model*

the outside and inside temperatures. The residential home never experienced the same inside and outside air temperatures when the heating was used and therefore never had the same conditions. With the first experimental model, this temperature difference was solved with an adjustment through subtracting the start temperatures. The offset model used this with an 'offset', which shifted the exponential curve by a time t_{offset}.

The two models were plotted together with the experimental data in figs. 6.28 and 6.29 and it was shown, that a temperature difference existed between both curves.

The offset approach as a curve fit revealed the temperatures, which were inferred from observation in Subsection 6.5.4 on page 109. Those two parameters were the outside temperature at the end of the heating and the temperature difference, which was found by the curve fit as $T_{out} = 11.47\,°C$ and $T_{diff} = 23.31\,°C$.

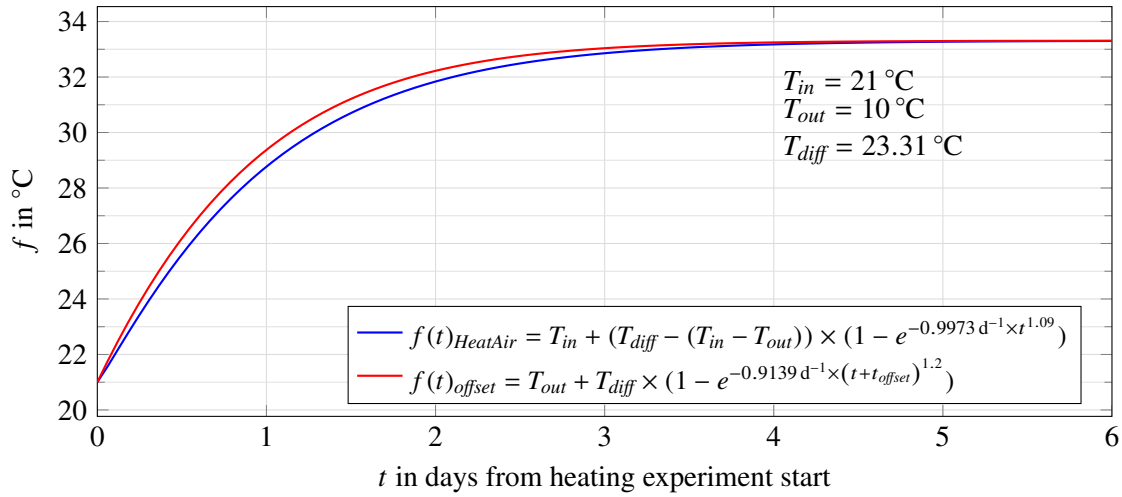


Figure 6.29: *Zoomed in landing air heating models*

6.7 Conclusion

This Chapter introduced the space conditioning experiments, conducted at the residential home. The proposed layout was to let the residential house cool down with a forced cooling, heat it for a certain amount of time and then turn off the heating to let the residential home cool again. Aspects not included in the analysis and creation of an experimental model for cooling and heating were outside influences that could have impacted the experimental data. This was also one of the reasons to select only first floor temperatures close to the centre of the house, where there were no direct influence of sunlight or from alternative heat sources. Models were created for the temperature data from the landing air and core temperatures. A stretched/compressed exponential was used to describe the experimental temperature data for the landing air and core.

Models were created to capture the uniqueness of the residential home, mainly by the use of exponential models and modelling those to the cooling and heating part of the experiment. The results were conclusive in terms of finding parameters fitting the experimental data.

Experiments were conducted in a laboratory space conditioning setup, where a simulation represented an enclosed space with integrated heating. The analysis and models concurred with findings that were observed in the experimental data: a consistent temperature difference between the final inside and outside temperature during heating. A similar parameter was found for the cooling, where it described a temperature difference between inside and outside temperatures, used for short term predictions. The knowledge from the laboratory space conditioning experiments was applied to the exponential models and turned the experimental models into models that could be used outside of the scope of the experiments for general predictions.

The models described at the end of this Chapter were used as a stepping stone for a second model described in Chapter 7, incorporating the unique characteristics of the residential home in parameters of an exponential function.

Chapter 7 Improved models

In this Chapter improved prediction models for the landing air temperature are presented. The improved models incorporated results from the laboratory space conditioning experiments and space conditioning experiments conducted in the residential home, described in Chapter 6. The Chapter begins by presenting a different approach for modelling. Improved models from Chapter 6 are presented with results of application of these models to night time data. The static inputs to the improved model were replaced with dynamic inputs and the improved models were benchmarked against night time data. At the end of this Chapter, a new system layout is presented, integrating the new model in a space conditioning application with a theoretical scenario describing the process and installation of such a system.

7.1 Probability model

The initial models presented in Chapter 5 used multiple sensors as inputs. The sources included: web scraped data, physical sensors and forecast data. It was reasoned that the use of disparate sources would increase the performance of landing air prediction models. An optimal model was not found and the introduction of an additional input could further increase or decrease the prediction performance, which was shown in Chapter 5. Therefore a different approach for creating models was proposed that used one input initially and introduced additional inputs one by one. Due to time constraints, only one input was selected and is presented in this Section.

The Golden dataset vacancy data was used and a statistical and visual analysis was performed to establish if a probability model could be created. The inputs to the probability model were the past data points of the landing air temperature and the output was the prediction of the landing air temperature.

Initially the new model was set up with the landing air temperature and the number of successive repetitions of the same temperature value. For example, when the last three recorded temperature values were 21.0625 °C, the probability for a repetition of this value was calculated by observing the next data point. The number of successive data points, the quantity of the successions and the percentage of a repeated temperature value can be seen in fig. 7.1. The variable K was the number of repetitions of the same temperature value. The value $K + 1$ represented the number K of successive repetitions and one additional repeat value. The percentage was calculated by dividing the total number of successive repetitions K by $K + 1$. The data points for the landing air were recorded at 15 min intervals. The plot shows, that temperature predictions with this probability model were correct in two of three cases for up to three successive repetitions of the same temperature value. For example, when a temperature value of 21 °C was recorded twice in succession, the probability

for a repetition of this temperature value was 66.36%. The rising percentage for eight to ten repetitions was observed because only nine vacancy datasets were recorded with this high number of repetitions.

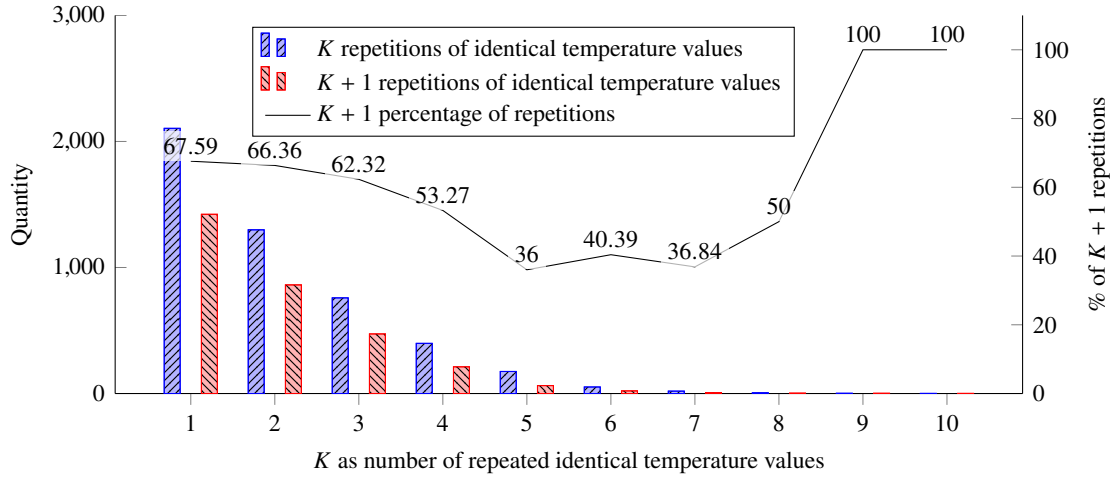


Figure 7.1: *Temperature value repetitions of the landing air temperature for the Golden dataset vacancies*

The next modelling attempt to create a probability model was set up with the landing air temperature gradients and repetitions of temperature gradients. For example, if the first temperature value was 21.0625 °C and the following temperature value was 21.125 °C, then the gradient was 0.0625 °C. A successful repetition was observed if the next temperature gradient value was 0.0625 °C. The resulting graph with the quantity of repetitions for K, $K + 1$, and percentages of the repeated gradient for $K + 1$ can be seen in fig. 7.2. The plot shows, that for two and three repetitions of the same gradient, 50% of the time the same gradient occurred again.

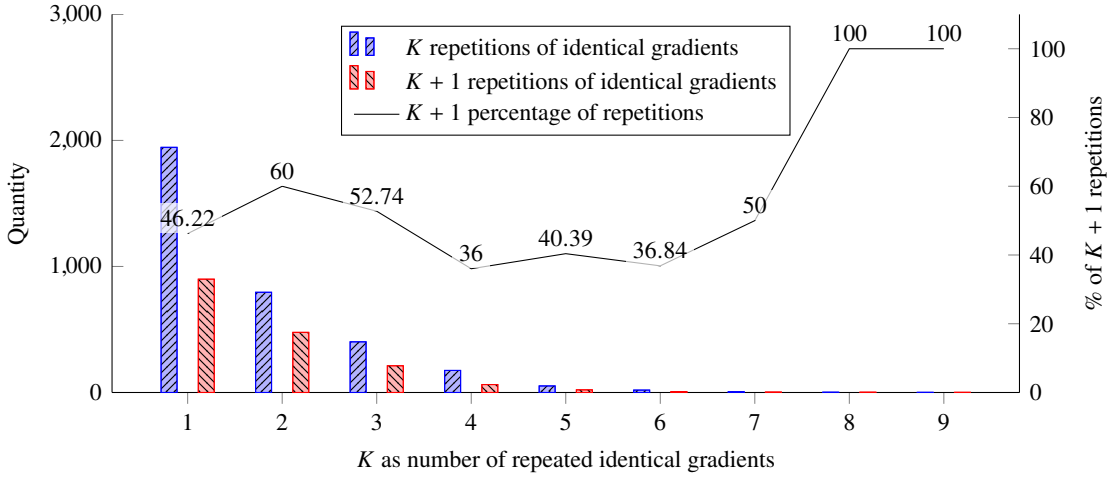


Figure 7.2: *Temperature gradient repetitions of the landing air temperature for the Golden dataset vacancies*

The probability model results showed, that short term predictions of 15 min yielded reliable results. However the time limitations for this model made it unsuitable for long term predictions.

7.2 Stretched exponential models and night time data

The vacancy data selected within the Golden dataset had dynamic inputs, which were not addressed with the stretched exponential models from the experiments. Therefore night time data was selected, as it showed consistent trending during night time cooling and during short heating phases in the morning. The consistency of the night time data enabled the application of the experimental stretched exponential models, described in Chapter 6.

The model was described in Subsection 6.5.4 on page 109 and used as inputs the landing air and the outside air temperature at the point when the heating was turned on or off. A plot of one of the sets can be seen in fig. 7.3, where the prediction data was plotted with the original night time data, and a simplified block diagram of the model can be seen in fig. 7.4.

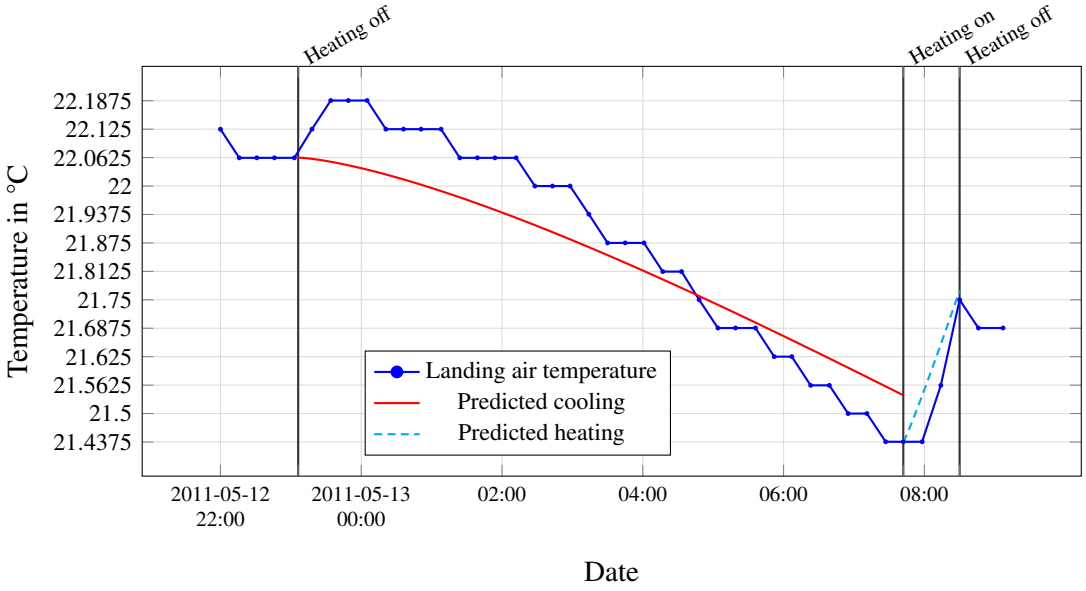


Figure 7.3: *Night time data with applied stretched exponential models for cooling and heating from the space conditioning experiments*

Table 7.1: *Stretched exponential models applied to data from the Golden dataset night time*

Set name/model	%MASE<1	$\overline{R^2}$	$\overline{\text{MAE}}$/°C
Night/cooling	76	0.45	0.12
Morning/heating	90	0.61	0.12

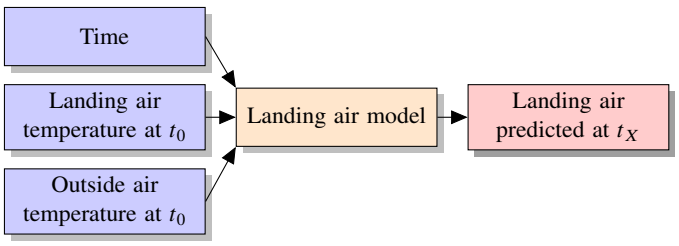


Figure 7.4: *Landing air prediction models for heating and cooling*

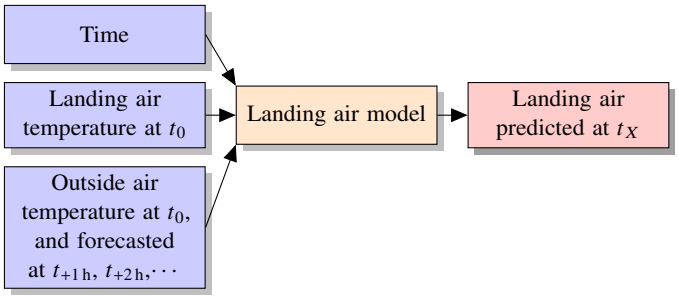


Figure 7.5: *Improved model with dynamic T_{out} input*

7.2.1 Dynamic inputs

The stretched exponential model applied to the Golden dataset night time data showed a qualitatively good fit. This dataset included data from all seasons except winter. The model was applied and validated against winter data from the After Experiment dataset and the Golden Extended dataset to confirm and compare results.

The stretched exponential model had temperatures T_{out} and T_{in} as inputs at the start of the prediction, time t_0. Several modifications were considered to make T_{out} and T_{diff} dynamic or include past and forecasted inputs. In the following paragraphs the proposed changes are presented and the results compared.

7.2.1.1 Dynamic T_{out}

The static input T_{out} of the outside temperature at the start time t_0 was replaced with hourly forecasts. The forecasted input took into account a change of the outside air temperature for each following hour. For example if a prediction between six and seven hours was made, the input would be a T_{out} that was six hours ahead and for a five to six hour prediction, the input was a T_{out} of a five hour forecast. A block diagram of this model can be seen in fig. 7.5.

7.2.1.2 Lookup table for T_{diff}

The models did not account for seasonal changes of outside air temperatures and used a static temperature difference of 6 °C to the inside air temperature. The seasonal changes were incorporated by picking different T_{diff} values for different seasons. The formula for cooling was then changed to implement the new parameter as follows:

$$f(t)_{CoolAir} = \left(T_{out} + T_{diff}\right) + \left(T_{in} - \left(T_{out} + T_{diff}\right)\right) \times e^{-0.864\,\mathrm{d}^{-1} \times t^{1.418}} \tag{7.1}$$

A block diagram showing the inputs to this model can be seen in fig. 7.6 on the next page. The model input T_{diff} referred to both the lookup table and the calculated values in Subsubsection 7.2.1.3.

The lookup table used for this dynamic input was initially set up with following values:

- 'Winter' was defined as the months December, January and February, with a T_{diff} = 7 °C.
- 'Spring' was defined as the months March, April and May, with a T_{diff} = 6 °C.
- 'Summer' was defined as the months June, July and August with a T_{diff} = 5 °C.

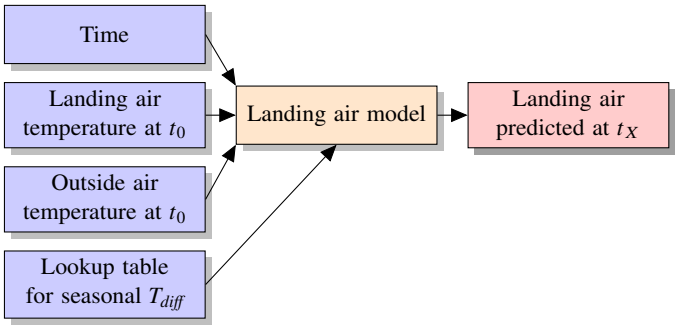


Figure 7.6: *Improved model with dynamic T_{diff} input*

- 'Autumn' was defined as the months September, October and November with a $T_{diff} = 6\,°C$.

The temperature value selection was based on the equilibrium analysis in Chapter 6, which revealed a mean value for $T_{diff} = 6\,°C$ during the mid-seasons and for the winter and summer seasons a change of $\pm 1\,°C$ was selected for the initial lookup table.

7.2.1.3 Calculated T_{diff} from past and forecasted data

Similar to the previous approach, T_{diff} was used as a dynamic input, dependant on previously observed or forecasted data. When the parameter was calculated from previous data, the previous day's cooling data was taken and the actual T_{diff} calculated with the following formula, which was derived from eq. (7.1):

$$T_{diff(\mathrm{prv})} = \frac{T_{end(\mathrm{prv})} - T_{out(\mathrm{prv})} + \left(T_{out(\mathrm{prv})} - T_{in(\mathrm{prv})}\right) \times e^{-0.864\,\mathrm{d}^{-1} \times t_{end(\mathrm{prv})}^{1.418}}}{1 - e^{-0.864\,\mathrm{d}^{-1} \times t_{end(\mathrm{prv})}^{1.418}}} \tag{7.2}$$

The calculated $T_{diff(\mathrm{prv})}$ was used as input to the improved model, as can be seen in fig. 7.6.

The variables used in eq. (7.2) were:

$T_{end(\mathrm{prv})}$ was the landing air end temperature from the previous day, recorded at $t_{end(\mathrm{prv})}$.

$T_{in(\mathrm{prv})}$ was the previous day's landing air start temperature.

$T_{out(\mathrm{prv})}$ was the previous day's outside air start temperature.

$t_{end(\mathrm{prv})}$ was the time in days from the start of the previous day's data collection until the end. For example 0.0416 for one hour.

If the previous day's data was not available, the lookup table temperature $T_{diff(\mathrm{lk})}$ was used instead.

If a forecast for the outside air temperature was used to predict the next day's $T_{diff(\mathrm{nxt})}$, an additional value was calculated based on values derived from previous day's data.

$$T_{end(\mathrm{nxt})} = T_{in} - \frac{T_{in(\mathrm{prv})} - T_{end(\mathrm{prv})}}{T_{end(\mathrm{prv})}} T_{in} \tag{7.3}$$

The values were then used in the following formula:

$$T_{diff(\mathrm{nxt})} = \frac{T_{end(\mathrm{nxt})} - T_{out(\mathrm{nxt})} + (T_{out(\mathrm{nxt})} - T_{in}) \times e^{-0.864\,\mathrm{d}^{-1} \times t_{end(\mathrm{prv})}^{1.418}}}{1 - e^{-0.864\,\mathrm{d}^{-1} \times t_{end(\mathrm{prv})}^{1.418}}} \tag{7.4}$$

Close-by days were observed and found to have similar temperature ranges and therefore it was generalised that a next day's data was similar to the previous day. The next day's data was then used in the same temperature range as the last day's data. The terms used for the calculation of a forecast value of $T_{diff(\mathrm{nxt})}$ were as follows:

$T_{end(\mathrm{nxt})}$ was the landing air end temperature calculated from previous day's data.

T_{in} was the current day's start temperature for the landing air.

$T_{out(\mathrm{nxt})}$ was the next day's outside air at t_{+24h}, taken from the forecasted outside air temperature.

$t_{end(\mathrm{prv})}$ was the time in days from the start of the previous day's data collection.

If more than one of the T_{diff} values was calculated, an average value was used instead.

7.3 Benchmarking results from dynamic input models

The benchmarking results for the models with varying dynamic inputs can be seen in table 7.2. The three measures, Mean Absolute Scaled Error (MASE), R^2 and Mean Absolute Error (MAE) were used to compare the model performances. The best performance values were highlighted in grey in each table. The MASE values were condensed in a new benchmarking factor '%MASE<1' which was the percentage of sets of better predictions than a naïve forecast compared to the total number of sets in a dataset. R^2 and MAE were averaged across all sets. Higher values for R^2 and %MASE < 1 were better and lower values better for MAE.

Table 7.2: *Dynamic inputs model benchmarking applied to night time data*

	Golden dataset			Golden Extended			After Experiment		
Model	%MASE<1	$\overline{R^2}$	$\overline{\mathrm{MAE}}$/°C	%MASE<1	$\overline{R^2}$	$\overline{\mathrm{MAE}}$/°C	%MASE<1	$\overline{R^2}$	$\overline{\mathrm{MAE}}$/°C
normal	79	0.366	0.126	89	0.538	0.149	90	0.483	0.202
$T_{out(\mathrm{fc})}$	73	0.261	0.140	86	0.426	0.167	91	0.445	0.192
$T_{diff(\mathrm{lk})}$	79	0.429	0.123	89	0.574	0.152	92	0.505	0.210
$T_{diff(\mathrm{prv})}$	82	0.435	0.134	90	0.579	0.162	94	0.573	0.173
$T_{diff(\mathrm{nxt})}$	69	0.228	0.181	82	0.358	0.232	90	0.422	0.237
$T_{diff(\mathrm{prv+nxt})}$	79	0.385	0.136	88	0.517	0.178	94	0.539	0.189
$T_{diff(\mathrm{prv+lk})}$	83	0.491	0.117	91	0.614	0.149	94	0.581	0.177
$T_{diff(\mathrm{nxt+lk})}$	78	0.400	0.131	88	0.520	0.174	94	0.521	0.202
$T_{diff(\mathrm{prv+nxt+lk})}$	83	0.461	0.122	91	0.576	0.161	95	0.568	0.183
$T_{out(\mathrm{fc})}+T_{diff(\mathrm{lk})}$	69	0.199	0.153	83	0.420	0.171	92	0.483	0.191
$T_{out(\mathrm{fc})}+T_{diff(\mathrm{prv})}$	71	0.255	0.167	84	0.447	0.185	94	0.492	0.177
$T_{out(\mathrm{fc})}+T_{diff(\mathrm{nxt})}$	56	0.058	0.216	77	0.268	0.239	87	0.359	0.231
$T_{out(\mathrm{fc})}+T_{diff(\mathrm{prv+nxt})}$	68	0.177	0.177	74	0.390	0.194	92	0.461	0.187
$T_{out(\mathrm{fc})}+T_{diff(\mathrm{prv+lk})}$	71	0.227	0.157	85	0.441	0.174	96	0.520	0.173
$T_{out(\mathrm{fc})}+T_{diff(\mathrm{nxt+lk})}$	68	0.182	0.170	88	0.392	0.178	92	0.478	0.191
$T_{out(\mathrm{fc})}+T_{diff(\mathrm{prv+nxt+lk})}$	70	0.208	0.165	84	0.423	0.180	93	0.502	0.177

The results showed that models with dynamic inputs improved their performance. The best performing models applied to night time for the Golden and Golden Extended dataset included

the dynamic input $T_{diff(\text{prv+lk})}$, yielding the highest %MASE<1, the best average R^2 rate and the smallest average MAE. To compare the models generated from the Golden and Golden Extended dataset, the After Experiment dataset was used, but no best model was identified for all three datasets. The %MASE<1 of all models applied to the After Experiment datasets was above 87%, showing a high success rate for predictions of night time cooling. The monthly distribution for the validation set can be seen in table 7.3 and it shows that the Golden dataset included fewer sets during colder weather conditions. The Golden Extended dataset had a wider spread of data across all seasons. The After Experiment dataset covered all seasons as well, but as it included data from a subsequent year, different results were obtained.

Table 7.3: *Number of sets for night time cooling with a minimum of one hour length*

Month	Golden	Golden Ext.	After Exp.
January	-	27	27
February	-	27	21
March	2	27	27
April	26	26	27
May	23	23	21
June	10	10	16
July	12	12	8
August	5	5	2
September	19	19	19
October	10	10	9
November	-	4	25
December	-	20	29
Σ	107	210	231

7.3.1 Improved T_{diff} for lookup tables

The $T_{diff(\text{lk})}$ took into account seasonal changes. A further step in improving the choices of temperatures values was carried out. Average temperatures recorded in the area of Emsworth can be seen in fig. 7.7 on the following page for the period from 1981-2010 (The Met Office, 2013a). The plot shows the monthly temperature distribution, which resembled a sine wave. On this plot, it can be seen that the seasons defined in Subsubsection 7.2.1.2 on page 117 did not have common temperatures. Therefore the months with same temperatures were used as a guideline. The monthly pairs and the selected temperature values for $T_{diff(\text{lk})}$ can be seen in table 7.4.

Table 7.4: *Seasonal $T_{diff(lk)}$ temperature values for each month*

Months	T_{diff}
January and February	8 °C
March and December	7 °C
April and November	6 °C
May and October	6 °C
June and September	5 °C
July and August	4 °C

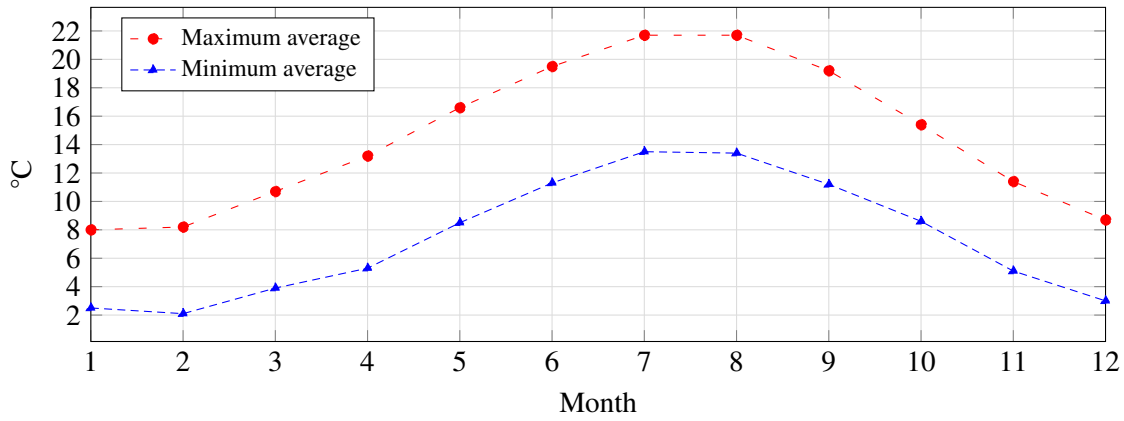


Figure 7.7: *Monthly average for maximum and minimum temperatures in the Emsworth area for the period 1981-2010*

The improved model was then validated against night time cooling data from Golden Extended and After Experiment datasets. The results can be seen in table 7.5. The Golden dataset was not included, as it was a subset of the Golden Extended dataset.

Table 7.5: *Improved model validation for cooling with different $T_{diff(lk)}$*

	Golden Extended			After Experiment		
Method	%MASE<1	$\overline{R^2}$	$\overline{\text{MAE}}$/°C	%MASE<1	$\overline{R^2}$	$\overline{\text{MAE}}$/°C
$T_{diff(\text{lk})}$	90	0.609	0.160	93	0.523	0.211
$T_{diff(\text{prv+lk})}$	92	0.634	0.154	94	0.590	0.179
$T_{diff(\text{nxt+lk})}$	89	0.532	0.178	94	0.528	0.204
$T_{diff(\text{prv+nxt+lk})}$	91	0.590	0.164	95	0.568	0.184
$T_{out(\text{fc})}+T_{diff(\text{lk})}$	83	0.442	0.178	94	0.499	0.191
$T_{out(\text{fc})}+T_{diff(\text{prv+lk})}$	84	0.442	0.181	95	0.516	0.175
$T_{out(\text{fc})}+T_{diff(\text{nxt+lk})}$	82	0.386	0.192	92	0.475	0.193
$T_{out(\text{fc})}+T_{diff(\text{prv+nxt+lk})}$	83	0.412	0.185	93	0.499	0.179

The results showed that for a model including $T_{diff(\text{prv+lk})}$ as input, it performed marginally better for the Golden Extended dataset. The After Experiment dataset results including $T_{diff(\text{prv+lk})}$ were slightly worse. The overall changes to the benchmarking results compared to table 7.2 on page 119 were marginally better for models including only $T_{diff(\text{lk})}$ and models including $T_{out(\text{fc})}$ performed poorer.

To find the best $T_{diff(\text{lk})}$ seasonal profile for a dataset, a range of temperature values were tested. The temperature ranges included values between 3 °C and 13 °C, with a resolution of 1 °C. The best fit of seasonal profile values for $T_{diff(\text{lk})}$ were identified with the highest %MASE<1 and the highest average R^2. Table 7.6 shows the initially selected values, taken from table 7.4 on page 120 and the best fit of values obtained for the Golden Extended and After Experiment dataset. It can be seen from the values, that the choices of $T_{diff(\text{lk})}$, shown in table 7.4, were close to the best fitting values found for the Golden dataset, however the values were consistently 1 °C lower for the After Experiment dataset. The benchmarking results for each dataset with the best tuned $T_{diff(\text{lk})}$ values can be seen in table 7.7 on the next page.

Table 7.6: *Initial and best fit of $T_{diff(lk)}$ values for datasets*

	Month											
Seasonal temperature difference profiles	1	2	3	4	5	6	7	8	9	10	11	12
$T_{diff(lk)}$/°C initial	8	8	7	6	6	5	4	4	5	6	6	7
$T_{diff(lk)}$/°C Golden Extended	7	7	7	5	6	5	4	4	5	6	5	7
$T_{diff(lk)}$/°C After Experiment	7	7	6	5	5	4	4	4	4	5	5	6

Table 7.7: *Model benchmarks for cooling predictions with tuned $T_{diff(lk)}$ values for the Golden Extended and for the After Experiment dataset*

	Golden Extended			After Experiment		
Model	%MASE<1	$\overline{R^2}$	$\overline{MAE}$/°C	%MASE<1	$\overline{R^2}$	$\overline{MAE}$/°C
$T_{diff(lk)}$	91	0.624	0.148	95	0.565	0.185
$T_{diff(prv+lk)}$	92	0.645	0.147	96	0.607	0.164
$T_{diff(nxt+lk)}$	88	0.539	0.174	93	0.530	0.194
$T_{diff(prv+nxt+lk)}$	91	0.594	0.160	94	0.571	0.176
$T_{out(fc)}+T_{diff(lk)}$	82	0.415	0.180	94	0.439	0.191
$T_{out(fc)}+T_{diff(prv+lk)}$	84	0.429	0.181	93	0.476	0.177
$T_{out(fc)}+T_{diff(nxt+lk)}$	82	0.374	0.193	91	0.443	0.194
$T_{out(fc)}+T_{diff(prv+nxt+lk)}$	82	0.402	0.185	92	0.473	0.181

These results showed, that the models with the dynamic inputs $T_{diff(prv+lk)}$ had the best performance values. The $T_{diff(lk)}$ and combined $T_{diff(prv+lk)}$ showed improvements, but models including a forecasted $T_{out(fc)}$ performed worse.

This approach only worked because full datasets including future data were available. Instead a new method of calculating $T_{diff(lk)}$ was proposed, which would use averaged values of $\Delta T_{air} = max(T_{air}) - min(T_{air})$ for the inside air temperature to deduce the temperature difference between inside and outside, based on day-to-day data. When the values from table 7.6 were plotted together with ΔT_{air}, the curve shape indicated that the $T_{diff(lk)}$ could be calculated, as can be seen in fig. 7.8. The analysis and further investigation of the similarities was not carried out in this Dissertation. The marginal improvements achieved with these adjustments could lead to model predictions with success rates higher than 92% for %MASE<1. The static $T_{diff(lk)}$ models however showed performance values of over 90% and were recognized as sufficient to produce reliable predictions.

7.3.2 Heating model

The cooling model was used to predict times of cooling between night time and the following morning. An additional observed temperature change in the residential home was a heating phase in the morning, initiated by the heating being turned on. These sets were subjected to benchmarks, when the heating was on for a minimum of one hour. A minimum of one hour was selected to have sufficient data points recorded, which was compared to the predicted data. The T_{diff} for heating was static and therefore no dynamic values were calculated. The temperature values for $T_{out(fc)}$ were used for forecasted values of the outside temperature. Furthermore two different heating models were used to validate against the data: the experimental heating model and the offset heating model as described in Section 6.6 on page 110.

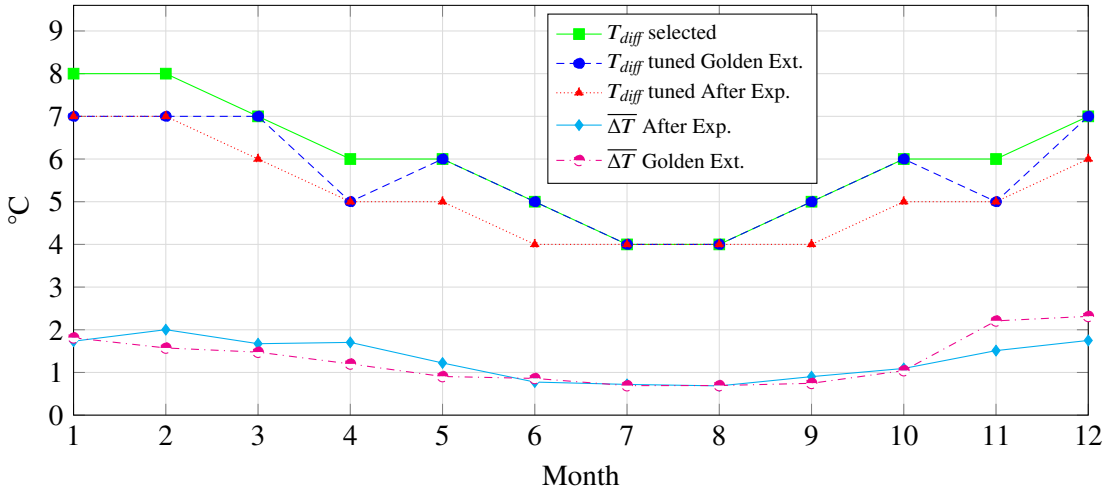


Figure 7.8: *Comparison of $T_{diff(lk)}$ and ΔT for different datasets*

Table 7.8: *Stretched exponential and improved models for heating applied to night time data*

	Golden Extended (81 sets)			After Experiment (143 sets)		
Model	%MASE<1	$\overline{R^2}$	$\overline{\text{MAE}}$/°C	%MASE<1	$\overline{R^2}$	$\overline{\text{MAE}}$/°C
normal	88	0.631	0.115	81	0.535	0.164
$T_{out(\text{fc})}$	89	0.639	0.114	81	0.514	0.164
normal (offset model)	78	0.474	0.170	79	0.509	0.190
$T_{out(\text{fc})}$ (offset model)	80	0.475	0.170	79	0.503	0.189

The results show, that the dynamic input $T_{out(\text{fc})}$ had a positive impact when included in the model inputs. This result was in line with the cooling model and reiterated the importance of including forecast values for an improved prediction performance. The prediction model block diagram can be seen in fig. 7.5 on page 117.

7.3.3 Discussion

The improved models for cooling and heating showed qualitatively good performance values for landing air temperature predictions during night time and early morning hours. The selection of these times of day benefited from inhabitants dormancy and fewer dynamic influences (for example solar radiation) on the residential home.

The best temperature prediction models for the residential home were identified as:

$$f(t)_{cool} = (T_{out} + T_{diff(prv+lk)}) + (T_{in} - (T_{out} + T_{diff(prv+lk)})) \times e^{-0.864\,\mathrm{d}^{-1} \times t^{1.418}} \tag{7.5}$$

$$f(t)_{heat} = T_{in} + (T_{diff} - (T_{in} - T_{out(fc)})) \times (1 - e^{-0.997\,\mathrm{d}^{-1} \times t^{1.09}}) \tag{7.6}$$

The forecasted T_{out} was not beneficial to both the cooling and heating model, which can be seen in table 7.2 on page 119 and table 7.8 on page 123. The correlation analysis results of hourly forecasts in table 4.6 on page 54 compared local and forecasted temperatures. Short term forecasts were found to be the best fitting forecasts, which concurs with the model results. Due to the selection of night time data, prolonged cooling data was available. In contrast, only short periods of morning heating were available. The difference in available test data could explain the differing effects of T_{out} on the cooling and heating models.

The models had limitations that could be addressed in future work:

- Solar influences were not modelled.
- Other possible predictors to the model, such as additional sensors, were not considered.
- Additional sensors could be installed to monitor inhabitants behaviour to establish when they are asleep, rather than basing this information on general deductions.

7.4 Application of the improved model

The newly created models described in eqs. (7.5) and (7.6) were able to predict night time cooling and early morning heating, which addressed following use cases:

- When does the heating need to be turned off to reach temperature T_X at time t_{+X} (optimal stop)?
- When does the heating need to be turned on to reach temperature T_Y at time t_{+Y} (optimal start)?

A flowchart depicting integration and application of the novel non-linear prediction models can be seen in fig. 7.10 and a graph showing a result of the application of such a system with predicted and observed temperatures is shown in fig. 7.9. The proposed heating control system design could be scheduled to run after sensor data was collected to control the heating appliance.

The plot in fig. 7.9 describes the system behaviour when presented with example data. The red line plot was example temperature data T_{in}, corresponding to inside air temperature. The event times t_1 and t_2 were inhabitants' events, such as 'going out' in the morning and 'coming back' in the evening. The temperatures T_{t_1} and T_{t_2} correlated to the event times t_1 and t_2 connected

with respectively with a green dashed line. A red dot marked the crosspoint of the correlated times and events. For example: inhabitants went to bed at ten o'clock (t_1) in the evening and preferred a temperature of 18° (T_{t_1}). The program flowchart in fig. 7.10 was executed in regular intervals when sensor data was recorded, which can be seen in the plot as the step-like appearance of the temperature T_{in}. Each thin grey line corresponded to the execution of the program flowchart, connecting the input temperature and a corresponding predicted temperature at the event times t_1 and t_2. The first half of the plot shows the application of optimal stop, with a blue background, and the second half of the plot shows the application of optimal start of the heating appliance with an orange background. The optimal stop path of the flowchart calculated a predicted inside air temperature for the time t_1 and compared the result with the event temperature T_{t_1}. If the result was equal or higher than T_{t_1}, the heating appliance was turned off, shown as thick blue line. For the second half of the plot, the optimal start path was followed in the flowchart, applying the heating model. The thin grey lines connect the begin and end temperatures calculated for t_2. When the predicted inside air temperature was equal or smaller than T_{t_2}, the heating was turned off, marked with a second thick blue line.

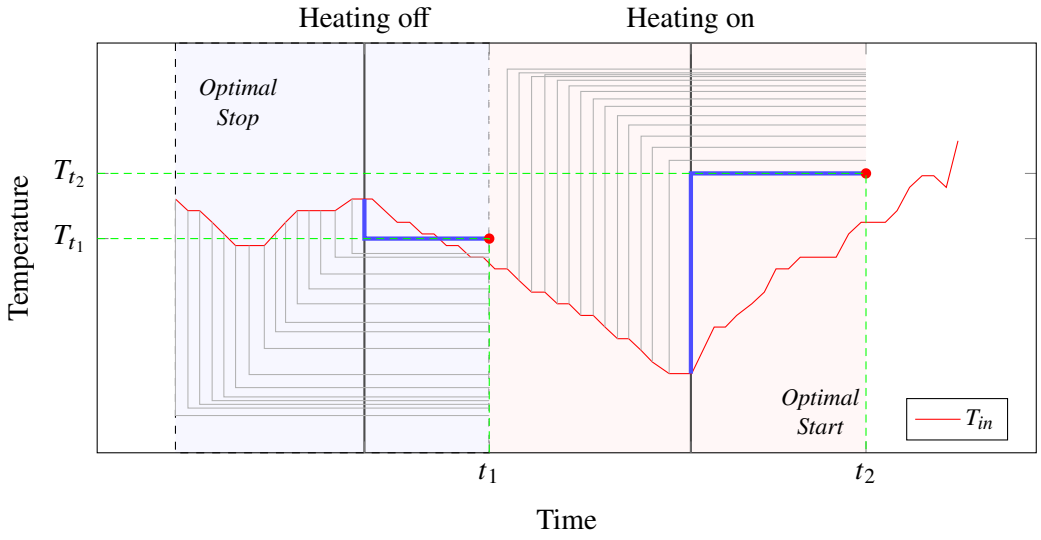


Figure 7.9: *New heating system applied to example data*

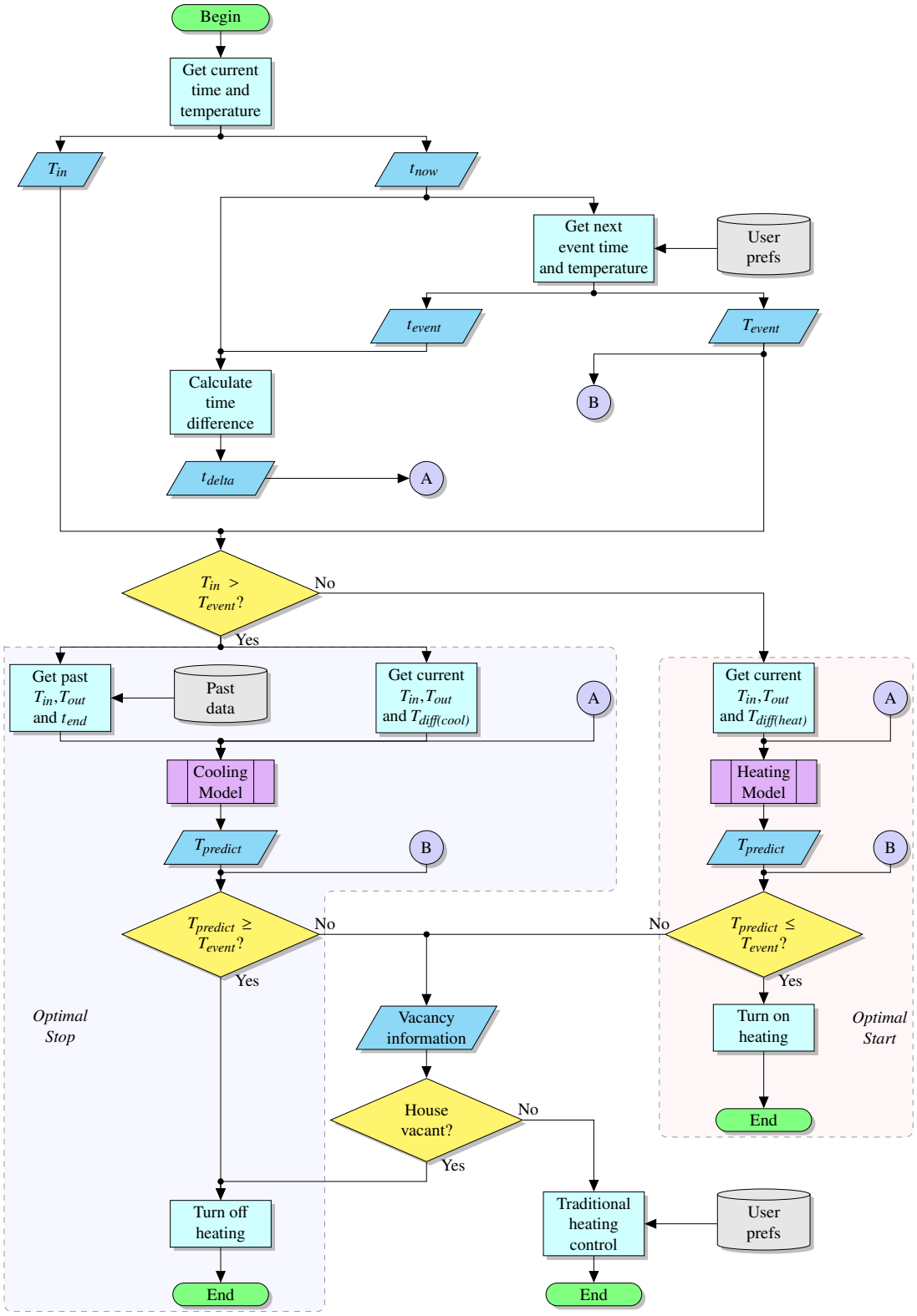


Figure 7.10: *New heating system design integrating air temperature prediction models*

7.5 Future system scenario

The final improved models described in this Chapter could be used in future systems, which could be realised in small computer systems, for example a Raspberry Pi[1]. The system would collect data and provide sufficient computing power to process real-time analysis. In addition to the network interfaces, the size was comparable to that of thermostats used in a residential home and therefore paved the way for an unobtrusive installation. The introduction of wireless network enabled routers would make wireless connections a viable option for a final wall mounted system without any wiring necessary except for a power supply.

A use case for the described systems is presented here, with an example of a residential home inhabitant's preferences and possible enhancements to the system.

7.5.1 User case study

A few days before Bob moved into his residential home, he installed the newly developed system 'X' in his residential home. The manual accompanying system 'X' asked the user to allow the system to collect data throughout the next couple of weeks. During that time, the collected data would incrementally adjust the default cooling and heating models to fit the collected data. After a few weeks, the system would have collected enough data to deduce times of vacancy and could carry out heating experiments, with prior permission of Bob. To further increase the prediction accuracy, longer times of vacancy, such as holidays were typed in by Bob to allow for longer vacancy experiments. This was only needed, if the short vacancies did not create reliable prediction models better than the naïve forecast in 50% of the time the prediction models were applied.

After moving in, Bob used the input device for system 'X' to select if temperatures were at a comfortable level. The input device was a knob with a red and a blue zone, indicating if the current temperatures was too warm or too cold. With the help of passive infrared and alarm system sensors integrated into system 'X', specific reoccurring events such as 'Dormancy', 'Awakening', 'Return to premises', and 'Weekend' were created. Taking into account the temperature preferences, the system created a table connecting preferred temperatures with reoccurring events. For example: in the evening, Bob preferred the ambient temperature to be slightly lower and in the morning and evenings he preferred even lower temperatures. During the day, no heating was necessary and turned on in time for Bob's return from work. The system would know when to turn the heating system 'on' and 'off' to reach certain temperatures with the knowledge of the parameters gained from the experiment. This scenario can be seen in fig. 7.11, depicting reoccurring events, preferred temperatures, heating states controlled by the system and real temperatures.

[1] http://www.raspberrypi.org

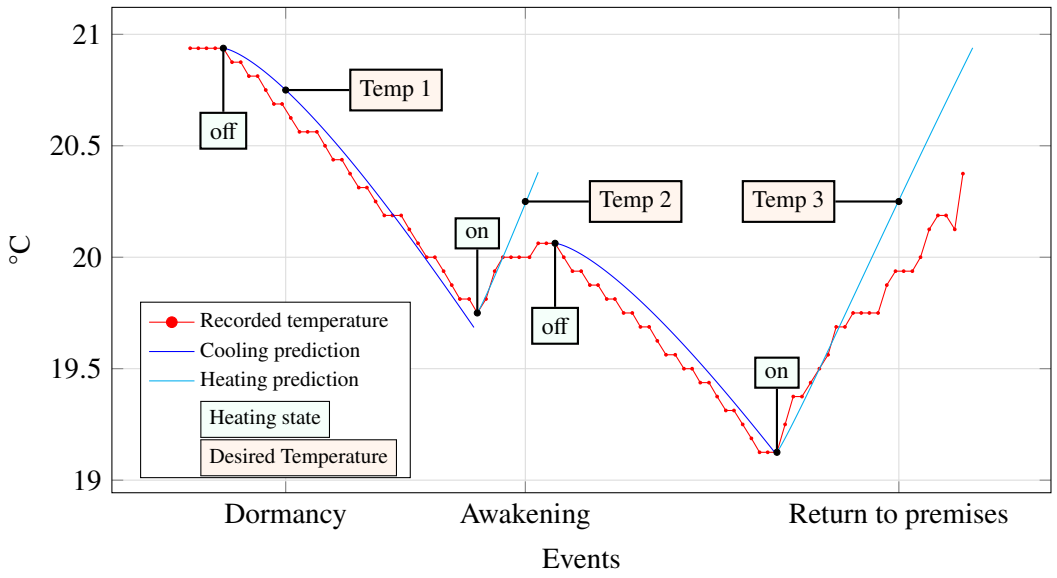


Figure 7.11: *Exemplary temperature data with cooling and heating prediction, showing events with desired temperatures and heating states*

7.6 Conclusion

The Chapter presented the improved model incorporating knowledge gained from the experiments carried out in the residential home. The non-linear stretched exponential models created in Chapter 6 were used with the following inputs: time, landing air temperature, outside air temperature, forecasted temperatures, and lookup tables for temperature differences. The predictions were made during night times to decrease the impact of dynamic inputs such as solar radiation and inhabitants. The cooling and heating predictions of the landing air were found to be better than a naïve forecast, which made this system perform better than the initial model, presented in Chapter 5. Including forecasts and knowledge from previous days increased prediction performance rates to over 90%. The results showed, that the use of stretched exponential models derived from experimental data showed a qualitatively good performance and were suitable for day-to-day prediction of temperatures in the residential home.

The newly developed improved models were able to predict temperature data for cooling and heating applications and were suitable to answer following questions:

- When does the heating need to be turned off to reach temperature T_X at time t_{+X}?
- When does the heating need to be turned on to reach temperature T_Y at time t_{+Y}?

The application of the prediction models would be in the area of optimal start/stop control, which has the potential of saving 5-10% of the heating energy (CIBSE, 2004).

Chapter 8 Conclusion

In this Chapter conclusions from the results are discussed, the contributions presented, and future work pointed out.

8.1 Summary of the research

The encompassing aim of the research was to create a prediction model for inside air temperatures in a vacant residential home that used environmental and experimental data from disparate sources.

The work was carried out in several stages. Initially a new system was created to collect data from various local and remote sensors. Then data mining methodology was applied to prepare the data for analysis. In a subsequent step initial prediction models were created. Experiments carried out in a residential home and in a laboratory facilitated the creation of a new experimental prediction model, which could be used in general temperature predictions in the residential home. A final step produced an improved version from the experimental model and a novel design for a residential heating application was formulated, incorporating the new improved models.

8.2 Resolution of research aims and objectives

The specific objectives to reach the aim are summarised below. Beneath each of the objective is an assessment summary of the work completed.

- *To create a system that collects data from local, remote, and aggregated sensors.*
  A residential home had local environmental sensors installed at the beginning of the research and several additional sensors were installed throughout winter 2010 and spring 2011, described in Chapter 3 on page 21. Third party environmental sensors were accessed through websites and an on-site weather station was added to collect climate data at the residential home in October 2010. Sensors were combined to create aggregated sensors, such as vacancy and heating detection.

- *To create prediction models from historical data.*
  Raw sensor data was pre-processed prior to the creation of models. Data mining techniques were employed, which are described in Chapter 4 on page 36. The raw data contained outliers, which were removed and data was normalised for the initial modelling process. Initial prediction models were created and described in Chapter 5 on page 59. The prediction models were used to predict inside air and brick temperatures in a vacant unheated residential home. The inclusion of the vacancy and heating detection sensors marked the start of data collection in April 2011. The initial models were created with the collected historical

environmental data. Improvements to the initial model process included a new validation method and an iterative modelling approach. The more historical data was available for the creation of the iterative prediction model, the better the prediction performance in the beginning. However the historical environmental data was partitioned into datasets because of structural changes in the residential home. Each partitioned dataset included data of less than one year and had no repeated season. Furthermore the temperature differences during unheated vacancies were found to be small in comparison to daily temperature differences. This showed, that the prediction results were accurate over the limited temperature ranges encountered in unheated vacancies.

- *To conduct space conditioning experiments in a residential home and create temperature prediction models from the experimental data.*
  To conduct and analyse experiments on laboratory space conditioning experiments.
  Space conditioning experiments were conducted at the residential home and are described in Chapter 6 on page 87. It was hypothesised, that the residential home's response to a step input could help identify and create better prediction models. The step inputs consisted of continuous heating and continuous cooling of the residential home. Due to time constraints, the observed temperatures inside the residential home did not settle at final temperature values. Different modelling approaches were applied, arriving at exponential models describing the experimental data. Attempts to unify the heating and cooling parts of the experimental model were unsuccessful. Two experimental models were therefore created: one for heating and one for cooling. Further space conditioning experiments were conducted in a laboratory setup, where a wooden box and bricks were used to simulate the residential home. The experiments were carried out in an environmental chamber at different ambient temperatures. The box was subjected to cycles of heating on and heating off. The results from the laboratory space conditioning experiments were applied to the residential home's experimental models with parameters describing temperature differences. The new experimental prediction models including the parameters were able to be applied to temperature predictions in unheated vacant space in the residential home.

- *To create a new prediction model that can be used in heating control applications for residential homes.*
  A different approach was investigated where a temperature prediction model used a single input and a single output in a probability model. This prediction model yielded satisfactory results for short term predictions of up to 45 minutes, presented in Chapter 7 on page 114. This model was not further investigated, because its application was limited to short term predictions. The experimental prediction models were applied to night time data, where larger temperature differences were observed. Furthermore, night times were undisturbed by dynamic daytime influences. Static inputs of the experimental prediction models were replaced with dynamic inputs, incorporating forecasts and seasonal changes to create a new improved prediction model. The predicted temperatures had an average error margin of $\pm 0.12\,^{\circ}\mathrm{C}$. The improved cooling and heating models showed an improved performance

compared to the initial models and the experimental models, when applied to night time data for validation. The results showed, that models derived from experimental data showed a qualitatively good performance and were suitable for day-to-day prediction models in the residential home.

The research work concluded with the design of a heating system incorporating the new improved prediction models. The new heating control could be applied in the area of optimal start-stop, where the following questions were answered:

- When does the heating need to be turned off to reach temperature T_X at time t_X?
- When does the heating need to be turned on to reach temperature T_Y at time t_Y?

The analysis of experiments carried out in a residential home and the analysis of laboratory space conditioning experiments facilitated the creation of novel non-linear temperature prediction models. The novel prediction models were integrated into a new design for a residential heating system.

8.3 Key research contributions

The aims and objectives set out at the beginning of this Dissertation were completed and the following contributions were achieved:

- Creation of
  - novel linear regression models that used environmental historical data recorded from local and remote sensors to predict inside air temperature in a residential home.
  - a new non-linear heating model to predict inside air temperatures.
  - a new non-linear cooling model to predict inside air temperatures.
  - novel linear and non-linear prediction models that included web-scraped weather forecast data.
  - a new method to define a static temperature differences input parameter for inside air and wall temperature prediction models.
  - new methods to define dynamic temperature differences input parameters for inside air temperature prediction models.
  - a revised novel non-linear model to predict inside air temperatures that applied knowledge from laboratory space conditioning experiments to prediction models created from experimental data.
  - new software system to collect data from local and remote sources for the creation of prediction models.
  - a new non-linear approximate heating model for a laboratory space conditioning experiment.
  - a new non-linear approximate cooling model for a laboratory space conditioning experiment.

  - a new iterative linear model to predict inside air temperatures that used accumulated environmental data.
  - a new linear model to predict inside air temperatures with different inputs for different prediction lengths.

- Application of
  - a step-input to a residential heating system to create models as an input for a system that used disparate sources of information to predict temperature in a vacant space.
  - non-linear Kohlrausch exponential functions to create novel prediction models.

- Design of a new residential optimal start-stop heating system that incorporates novel non-linear models to predict inside air temperatures.

The key contribution was the design of a new residential heating control system that applied the novel non-linear prediction models for inside air temperatures based on experimental and historical environmental data.

8.4 Suggestions for future work

A list of suggestions to further the research presented in this Dissertation as follows:

- The research was focussed on one residential home and the application and analysis should be widened to several residential homes of different building structures.
- The new heating control design should be applied and the efficiency should be analysed and compared to traditional heating methods.
- The creation, design and programming of a single system should be considered, incorporating data collection from disparate sources, analysis, and control of the heating system, for example a RaspberryPi. Such a system would enable easy deployment to different residential homes and facilitate the use of a single programming language.
- An investigation of similarities and dissimilarities between parameters from space conditioning experiments on residential homes and those conducted in laboratory space conditioning criteria should be considered, to facilitate different methods of finding parameters for prediction models.
- During this research, two sensors were mainly investigated. More sensors inside a residential home should be analysed, modelled, and predicted.
- Night time data should be analysed and different models for night time and day time data should be considered.
- Thermal comfort should be included as a goal for a new heating control.
- The collection and usage of inhabitants' thermal comfort should be investigated further.

- The space conditioning experiment carried out on the residential home should be repeated in all seasons to widen the applicability of the model.
- The space conditioning experiment should be conducted as long as necessary to observe final temperature values.
- The 1wire networked sensor data reads would benefit from splitting up into sub networks with different master devices to decrease the time to read all sensors' data.
- All sensor data read intervals should be decreased to intervals of less than or equal to five minutes, as some sensor data was only refreshed every 15 or 30 minutes. The missing sensor data was interpolated where necessary, which could be avoided with a general read interval for all sensors.

8.5 Thesis conclusion

The new non-linear temperature prediction models were of a sufficiently high order to forecast inside air temperatures in a residential home and sufficiently simple to be understood by future specialist operators and inhabitants. This was made possible by the use of basic inputs into the prediction models. Parameters describing the structure were created by conducting space conditioning experiments in a residential home. The experimental setup could assist future applications, where inhabitants could install and use a heating control system in line with the proposed novel design.

The limitations of the new models were pointed out and further research is necessary to address them. The new models were incorporated into a new design of a heating control application, which could be used to verify and apply the new models. Possible applications in other residential homes should be considered in addition to widening the scope to include day time data and dynamic inputs.

It was shown in this Dissertation, that results from a simplified prediction model for inside air temperature predicted the observed behaviour and could be considered for further investigation and application in the space heating and Smart Home area.

Bibliography

Achterbosch, G. G. J., de Jong, P. P. G., Krist-Spit, C. E., van der Meulen, S. F. & Verberne, J. (1985). The development of a convenient thermal dynamic building model. *Energy and Buildings*, *8*(3), 183–196.

Anderssen, R. S., Husain, S. A. & Loy, R. J. (2004). The kohlrausch function: properties and applications. *Anziam Journal*, *45*, C800–C816.

ASHRAE. (2010). *ANSI/ASHRAE Standard 55-2010: thermal environmental conditions for human occupancy*. American Society of Heating, Refrigerating and Air-Conditioning Engineers.

Augusto, J. C. & Nugent, C. D. (Eds.). (2006). *Designing smart homes - the role of artificial intelligence*. Lecture Notes in Computer Science. Springer Berlin / Heidelberg.

Azzi, D., Loveday, D. L., Azad, A. K. M. & Virk, G. S. (1997). Multivariable modelling for building energy management. In *IEEE colloquium on modelling and simulation for thermal management (digest no. 1997/043)* (pp. 3–1). IET.

Balan, R., Stan, S.-D. & Lapusan, C. (2009). A model based predictive control algorithm for building temperature control. In *Proceedings of 3rd IEEE international conference on digital ecosystems and technologies, 2009. DEST '09* (pp. 540–545).

Balaras, C. A. (1996). The role of thermal mass on the cooling load of buildings. an overview of computational methods. *Energy and Buildings*, *24*(1), 1–10.

Belkin. (2013). WeMo - Smart Home Automation. Retrieved 2013-07-25, from http://www.belkin.com/us/wemo

Berglund, U. S. & Lundberg, B. H. (2000). Comfort control system incorporating weather forecast data and a method for operating such a system. US Patent.

Bertini, I., Ceravolo, F., Citterio, M., De Felice, M., Di Pietra, B., Margiotta, F., ... Puglisi, G. (2010). Ambient temperature modelling with soft computing techniques. *Solar Energy*, *84*(7), 1264–1272.

Bone, A., Murray, V., Myers, I., Dengel, A. & Crump, D. (2010). Will drivers for home energy efficiency harm occupant health? *Perspectives in Public Health*, *130*, 233–238.

Bonhomme, S., Campo, E., Esteve, D. & Guennec, J. (2007). An extended PROSAFE platform for elderly monitoring at home. In *29th annual international conference of the IEEE engineering in medicine and biology society, 2007. EMBS 2007* (pp. 4056–4059).

Braun, J. E. (2003). Load control using building thermal mass. *Transactions - American Societcy of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)*, *125*(3), 292–301.

Briggs, J., Curry, R. G. & Madge, B. (2002). Establishing an online telemedicine community in the uk. In *Proceedings of telemed '02: 9th international conference on telemedicine and telecare, london*.

BSI. (2005). *ISO 7730: ergonomics of the thermal environment-analytical determination and interpretation of thermal comfort using calculation of the pmv and ppd indices and local thermal comfort criteria*. The British Standards Institution. British Standards Institution.

BSI. (2007a). *BS EN 15217:2007 energy performance of buildings - methods for expressing energy performance and for energy certification of buildings*. British Standards Institution.

BSI. (2007b). *BS EN ISO 6946:2007 - building components and building elements - thermal resistance and thermal transmittance*. British Standards Institution.

BSI. (2008). *BS EN ISO 13790:2008: energy performance of buildings - calculation of energy use for space heating and cooling*. British Standards Institution.

BSI. (2012). *BS EN 15232:2012 energy performance of buildings — impact of building automation*. British Standards Institution.

Chalabi, Z. S., Bailey, B. J. & Wilkinson, D. J. (1996). A real-time optimal control algorithm for greenhouse heating. *Computers and Electronics in Agriculture*, *15*(1), 1–13.

Chan, M., Campo, E., Estève, D. & Fourniols, J.-Y. (2009). Smart homes – current features and future perspectives. *Maturitas*, *64*(2), 90–97.

Chan, M., Estève, D., Escriba, C. & Campo, E. (2008). A review of smart homes–present state and future challenges. *Computer Methods and Programs in Biomedicine*, *91*(1), 55–81.

Chan, M., Viard, T., Caillavet, M.-L. & Campo, E. (1995). Smart technology for the elderly and disabled users at home. In *The european context for assistive technology: proceedings of the 2nd tide congress, 26-28 april 1995, paris* (Vol. 1, p. 393). IOS Press.

Chapman, K. & McCartney, K. (2001, September). Smart homes for disabled people. In *The cutting edge 2001* (pp. 1–12). RICS Foundation, London.

Charles, K. E. (2003). *Fanger's thermal comfort and draught models*. NRC Institute for Research in Construction.

Cho, S. & Zaheer-uddin, M. (2003). Predictive control of intermittently operated radiant floor heating systems. *Energy Conversion and Management*, *44*(8), 1333–1342.

CIBSE. (2004). *Guide f: energy efficiency in buildings*. Chartered Institution of Building Services Engineers.

Cigler, J. (2013). *Model predictive control for buildings* (Doctoral Thesis, Czech Technical University in Prague).

Clinch, J. P. & Healy, J. D. (2001). Cost-benefit analysis of domestic energy efficiency. *Energy Policy*, *29*(2), 113–124.

Cook, D. J. (2007). Making sense of sensor data. *IEEE Pervasive Computing*, *6*(2), 105–108.

Cook, D. J., Youngblood, M., Heierman, E. O., Gopalratnam, K., Rao, S., Litvin, A. & Khawaja, F. (2003). MavHome: an agent-based smart home. In *Proceedings of the first IEEE international conference on pervasive computing and communications (PerCom'03)* (pp. 521–524).

Coyle, L. (2009). Construct. Retrieved 2009-11-29, from http://www.construct-infrastructure.org/publications/

Coyle, L., Neely, S., Rey, G., Stevenson, G., Sullivan, M., Dobson, S. & Nixon, P. (2006). Sensor fusion-based middleware for assisted living. In *Proceedings of the 4th international conference on smart homes and health telematics, ICOST'06* (pp. 281–288).

Coyle, L., Neely, S., Stevenson, G., Sullivan, M., Dobson, S. & Nixon, P. (2007). Sensor fusion-based middleware for smart homes. *International Journal of Assistive Robotics and Mechatronics (IJARM)*, *8*(2), 53–60.

Das, S. K., Cook, D. J., Battacharya, A., Heierman, E. O. & Lin, T.-Y. (2002). The role of prediction algorithms in the MavHome smart home architecture. *Wireless Communications, IEEE*, *9*(6), 77–84.

Davis Instruments. (2013). Vantage Pro 2 Weather Station. Retrieved 2013-05-22, from http://www.davisnet.com/weather/products/weather_product.asp?pnum=06152

De Dear, R. & Brager, G. S. (1998). Developing an adaptive model of thermal comfort and preference. *Indoor Environmental Quality (IEQ)*.

de Vergara, J. E. L., Villagrá, V. A., Fadón, C., González, J. M., Lozano, J. A. & Álvarez-Campana, M. (2008). An autonomic approach to offer services in OSGi-based home gateways. *Computer Communications*, *31*(13), 3049–3058.

Department of Energy and Climate Change. (2010). *Warm homes, greener homes: a strategy for household energy management*. HM Government - Department of Energy and Climate Change.

Desai, D. A. (2006). Physics through teaching lab – iv - newton's law of cooling. *Physics Education*, *23*(1), 51–54.

Dombayci, O. A. (2010). The prediction of heating energy consumption in a model house by using artificial neural networks in denizli-turkey. *Advances in Engineering Software*, *41*, 141–147.

Fadell, T. (2013). Nest - The Learning Thermostat. Retrieved 2013-07-23, from http://www.nest.com

Fanger, P. O. (1970). *Thermal comfort. analysis and applications in environmental engineering.* Copenhagen: Danish Technical Press.

Furfari, F., Sommaruga, L., Soria, C. & Fresco, R. (2004). DomoML: the definition of a standard markup for interoperability of human home interactions. In *Proceedings of the 2nd european union symposium on ambient intelligence* (pp. 41–44). Eindhoven, Netherlands: ACM.

Gegov, A. (2001). Intelligent modelling of the indoor climate in buildings by soft computing. *Mathware & Soft Computing*, (312-315).

Gellersen, H.-W., Schmidt, A. & Beigl, M. (2000). Adding some smartness to devices and everyday things. In *Third IEEE workshop on mobile computing systems and applications* (pp. 3–10).

Gibbons, M. (2011). *Heat transfer modelling in a domestic house* (Bachelor thesis, School of Engineering, University of Portsmouth).

Giorgetti, G., Gambi, E., Spinsante, S., Baldi, M., Morichetti, S. & Magnifico, I. (2008). An integrated solution for home automation. *2008 IEEE International Symposium on Consumer Electronics*, *1-2*, 157–160.

Government of Ireland. (2010). *Energy efficiency in traditional buildings*. The Department of the Environment, Heritage and Local Government.

Grünenfelder, W. J. (1985). The use of weather predictions and dynamic programming in the control of solar domestic hot water systems. *Mediterranian Electrotechnical Conference of IEEE Region 8*.

Guerra Santin, O., Itard, L. & Visscher, H. (2009). The effect of occupancy and building characteristics on energy use for space and water heating in dutch residential stock. *Energy and buildings*, *41*(11), 1223–1232.

Hancu, R. B. O., Stan, S., Lapusan, C. & Donca, R. (2010). Application of a model based predictive control algorithm for building temperature control. *Proceedings of the 3rd WSEAS International Conference on Energy Planning, Energy Saving, Environmental Education*, 97–101.

Harmo, P., Taipalus, T., Knuuttila, J., Vallet, J. & Halme, A. (2005). Needs and solutions - home automation and service robots for the elderly and disabled. In *IEEE/RSJ international conference on intelligent robots and systems IROS, 2005* (pp. 3201–3206).

Helal, S., Mann, W., El-Zabadani, H., King, J., Kaddoura, Y. & Jansen, E. (2005). The gator tech smart house: a programmable pervasive space. *Computer*, *38*(3), 50–60.

Helal, S., Mitra, S., Wong, J., Chang, C. & Mokhtari, M. (2008). *Smart homes and health telematics*. Springer, Berlin.

Hemphill, E. (2013). WigWag: Make stuff automatic. Retrieved 2013-07-25, from http://www.wigwag.com

Hong, X., Nugent, C. D., Mulvenna, M., McClean, S., Scotney, B. & Devlin, S. (2009). Evidential fusion of sensor data for activity recognition in smart homes. *Pervasive and Mobile Computing*, *5*(3), 236–252.

Hoyt, T., Schiavon, S., Piccioli, A., Moon, D. & Steinfeld, K. (2012). CBE Thermal Comfort Tool for ASHRAE-55. center for the built environment, university of california berkeley. Retrieved 2013-07-15, from http://cbe.berkeley.edu/comforttool/

Hudson, G. & Underwood, C. P. (1999). A simple building modelling procedure for MATLAB/SIMULINK. In *Proceedings of building simulation* (Vol. 99, *2*, pp. 777–83).

Hussain, S., Schaffner, S. & Moseychuck, D. (2009). Applications of wireless sensor networks and RFID in a smart home environment. In *Proceedings of seventh annual communication networks and services research conference, 2009. CNSR '09* (pp. 153–157).

Hyndman, R. J. (2006). Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, *4*, 43–46.

Hyndman, R. J. & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688.

Intille, S. S. (2002). Designing a home of the future. *Pervasive Computing, IEEE*, *1*(2), 76–82.

Jakkula, V. R., Cook, D. J. & Jain, G. (2007). Prediction models for a smart home based health care system. In *Proceedings of 21st international conference on advanced networking and applications workshops/symposia* (Vol. 7, 761–765).

Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.

Kastner, W., Neugschwandtner, G., Soucek, S. & Newman, H. M. (2005). Communication systems for building automation and control. *Proceedings of the IEEE*, *93*(6), 1178–1203.

KNX Association. (2013). Standard for home and building control: CENELEC EN 50090. Retrieved 2013-07-26, from http://www.knx.org

Kohlrausch, R. (1854). Theorie des elektrischen Rückstandes in der Leidner Flasche. *Annalen der Physik und Chemie*, *91*, 56–82, 179–214.

Kolokotroni, M., Perera, M., Azzi, D. & Virk, G. S. (2001). An investigation of passive ventilation cooling and control strategies for an educational building. *Applied Thermal Engineering*, *21*(2), 183–199.

Kosny, J., Petrie, T., Gawin, D., Childs, P., Desjarlais, A. & Christian, J. (2001). *Thermal mass - energy savings potential in residential buildings*. Buildings Technology Center, ORNL.

Li, K. W. (2012). *Modelling enclosure temperature* (Bachelor thesis, School of Engineering, University of Portsmouth). Bachelor thesis.

Lietzow, B., Dalheimer, M. & Walk, F. (2013). HexaBus - The kraken of home automation. Retrieved 2013-07-25, from http://hexabus.net/

Logan, B., Healey, J., Philipose, M., Tapia, E. & Intille, S. S. (2007). A Long-Term evaluation of sensing modalities for activity recognition. In *UbiComp 2007: ubiquitous computing* (pp. 483–500). Springer, Berlin.

Meier, C. (2010). *Richtig bauen: Bauphysik im Zwielicht - Probleme und Lösungen* (7th). Expert Verlag.

Mokhtari, M., Khalil, I., Bauchet, J., Zhang, D. & Nugent, C. D. (2009). *Ambient assistive health and wellness management in the heart of the city*. Springer, Berlin.

Moon, J. W. & Kim, J.-J. (2010). ANN-based thermal control models for residential buildings. *Building and Environment*, *45*(7), 1612–1625.

Moon, K.-D., Lee, Y.-H., Son, Y.-S. & Kim, C.-K. (2003). Universal home network middleware guaranteeing seamless interoperability among the heterogeneous home network middleware. *IEEE Transactions on Consumer Electronics*, *49*(3), 546–553.

Myatt, G. J. (2007). *Making sense of data: a practical guide to exploratory data analysis and data mining*. Wiley-Interscience.

Myatt, G. J. & Johnson, W. P. (2009). *Making sense of data ii: a practical guide to data visualization, advanced data mining, methods, and applications*. Making sense of data. John Wiley & Sons.

Nielsen, H. A. & Madsen, H. (2000). *Predicting the heat consumption in district heating systems using meteorological forecasts*. Lyngby, Denmark: Department of Mathematical Modelling, Technical University of Denmark.

Nugent, C. D., Finlay, D., Davies, R., Wang, H., Zheng, H., Hallberg, J., ... Mulvenna, M. (2007). homeML – an open standard for the exchange of data within smart environments. In *Pervasive computing for quality of life enhancement* (pp. 121–129). Springer.

Oldewurtel, F., Parisio, A., Jones, C. N., Gyalistras, D., Gwerder, M., Stauch, V., ... Morari, M. (2012). Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, *45*, 15–27.

Oldewurtel, F., Parisio, A., Jones, C., Morari, M., Gyalistras, D., Gwerder, M., ... Wirth, K. (2010). Energy efficient building climate control using stochastic model predictive control and weather predictions. In *American control conference (ACC), 2010* (pp. 5100–5105).

OpenStreetMap. (2013). Emsworth, United Kingdom. Retrieved 2013-05-22, from http://www.openstreetmap.org/?mlat=50.861481&mlon=-0.935876&zoom=10&layers=M/

Pahwa, A. & Brice, C. (1985). Modeling and system identification of residential air conditioning load. *IEEE Transactions on Power Apparatus and Systems*, *PAS-104*(6), 1418–1425.

Parnis, G. (2012). *Building thermal modelling using electric circuit simulation* (Doctoral dissertation, School of Photovoltaic and Renewable Energy Engineering, University of New South Wales, Sydney, Australia).

PassivSystems. (2013). PassivEnergy Wireless System. Retrieved 2013-07-24, from http://www.passivsystems.com/

Pham, V. T., Qiu, Q., Wai, A. A. P. & Biswas, J. (2007). Application of ultrasonic sensors in a smart environment. *Pervasive and Mobile Computing*, *3*(2), 180–207.

Privara, S., Siroky, J., Ferkl, L. & Cigler, J. (2011). Model predictive control of a building heating system: the first experience. *Energy and Buildings*, *43*(2–3), 564–572.

Rashidi, P., Youngblood, M., Cook, D. J. & Das, S. K. (2007). Inhabitant guidance of smart environments. In Jacko, JA (Ed.), *Proceedings of human-computer interaction, pt 2* (Vol. 4551, 910–919). Lecture Notes in Computer Science. 12th International Conference on Human-Computer Interaction (HCI International 2007), Beijing, PEOPLES R CHINA, JUL 22-27, 2007.

Ruano, A. E., Crispim, E. M., Conceição, E. Z. E. & Lúcio, M. M. J. R. (2006). Prediction of building's temperature using neural networks models. *Energy and Buildings*, *38*(6), 682–694.

Schwan, C. (2010). Die Temperierung - Eine neuartige Methode der Gebäudeheizung und der Trockenlegung durchnässter Wände, Kritik der konventionellen Heizmethode und der Energieeinsparverordnung EnEV. Retrieved 2013-07-24, from http://www.termosfassade.info/Dokumente/Temperierung.pdf

Schwan, C. (2013). Warum ist die EnEV vorgeschriebene U-Wert Berechnung falsch? Retrieved 2013-07-26, from http://www.termosfassade.info/Dokumente/Nr.40.pdf

Severne, R. (2011). *Heating efficiency and prediction* (Bachelor thesis, School of Engineering, University of Portsmouth).

Skrjanc, I., Zupancic, B., Furlan, B. & Krainer, A. (2001). Theoretical and experimental fuzzy modelling of building thermal dynamic response. *Building and Environment*, *36*(9), 1023–1038.

The Met Office. (2013a). Emsworth climate - averages table 1981-2010. Retrieved 2013-05-22, from http://www.metoffice.gov.uk/public/weather/climate/#?tab=climateTables

The Met Office. (2013b). How accurate are our public forecasts? Retrieved 2013-08-05, from http://www.metoffice.gov.uk/about-us/who/accuracy/forecasts

The Met Office. (2014). Regional climates: southern england. Retrieved 2014-04-13, from http://www.metoffice.gov.uk/climate/uk/so/

The UK Government. (2012). *Uk climate change risk assessment (CCRA): government report*. Department for Environment, Food & Rural Affairs.

The UK Government. (2013a). *Estimated impacts of energy and climate change policies on energy prices and bills*. Department of Energy & Climate Change.

The UK Government. (2013b). *The future of heating: a strategic framework for low carbon heat*. Department of Energy & Climate Change.

Tibbets, L. (2013). IFTTT: If This Then That. Retrieved 2013-07-25, from https://ifttt.com

Weiser, M. (1991). The computer for the 21st century. *Scientific American*, *265*(3), 94–104.

Weisstein, E. W. (2013a). 'Least Squares Fitting.' From MathWorld–A Wolfram Web Resource. Retrieved 2013-05-30, from http://mathworld.wolfram.com/LeastSquaresFitting.html

Weisstein, E. W. (2013b, June). 'Normal Distribution.' From MathWorld–A Wolfram Web Resource. Retrieved 2013-06-01, from http://mathworld.wolfram.com/NormalDistribution.html

Yang, I.-H. & Kim, K.-W. (2004). Prediction of the time of room air temperature descending for heating systems in buildings. *Building and Environment*, *39*(1), 19–29.

Zou, Y., Xie, C. & Lin, Z. (2007). Intelligent analysis system in time series of smart health home on-line monitoring data. In *Proceedings of IEEE international conference on control and automation, 2007. ICCA 2007* (pp. 1785–1790).

Appendix A Oversized tables and figures

Table A.1: *Full log of heating/cooling experiment conducted from 21st to 27th September 2011 in a residential home*

ID	Timestamp	Action/Comment	Inhab.	Doors	Windows
1	2011-10-21T10:00	Boiler turned off - cooling start	1	all open	all open
2	2011-10-21T15:30	-	3	all open	all open
3	2011-10-21T17:00	-	5	all open	all open
4	2011-10-21T18:00	Cooker on for 30 minutes	5	all open	all open
5	2011-10-21T18:30	All windows and doors shut	3	all shut	all shut
6	2011-10-21T22:00	-	5	all shut	all shut
7	2011-10-21T23:00	Bedtime	5	all shut	all shut
8	2011-10-22T11:00	All doors and windows open	3	all open	all open
9	2011-10-22T18:30	Downstairs doors and windows shut	0	downstairs shut	downstairs shut
10	2011-10-22T20:30	All doors and windows open	1	all open	all open
11	2011-10-22T23:45	Downstairs front door and windows shut	1	front door shut	front windows shut
12	2011-10-23T08:30	All doors open	1	all open	all open
13	2011-10-23T09:50	Downstairs doors and windows shut	1	front shut	downstairs shut
14	2011-10-23T12:40	doors and windows open	1	all open	all open
15	2011-10-23T14:00	Doors and windows shut	1	all shut facing outside	all shut
16	2011-10-23T14:15	heating start	1	all shut facing outside	all shut
17	2011-10-24T08:00	House empty	0	all shut facing outside	all shut
18	2011-10-25T14:15	Heating stop - cooling start	0	all shut facing outside	all shut
19	2011-10-27T08:00	Experiment ends - cooling stop	0	all shut facing outside	all shut

Appendix B Schematics

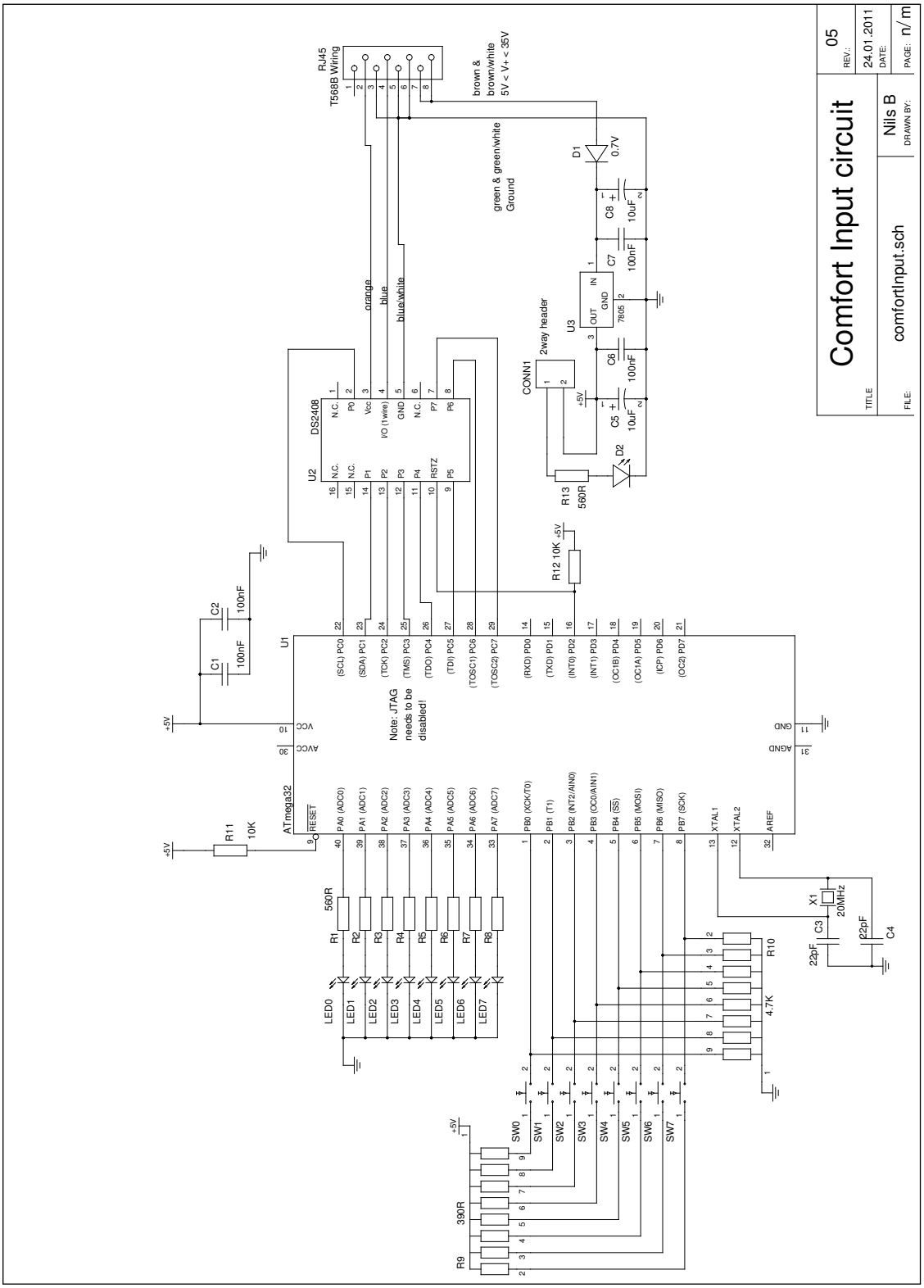


Figure B.1: *Comfort input box schematics*

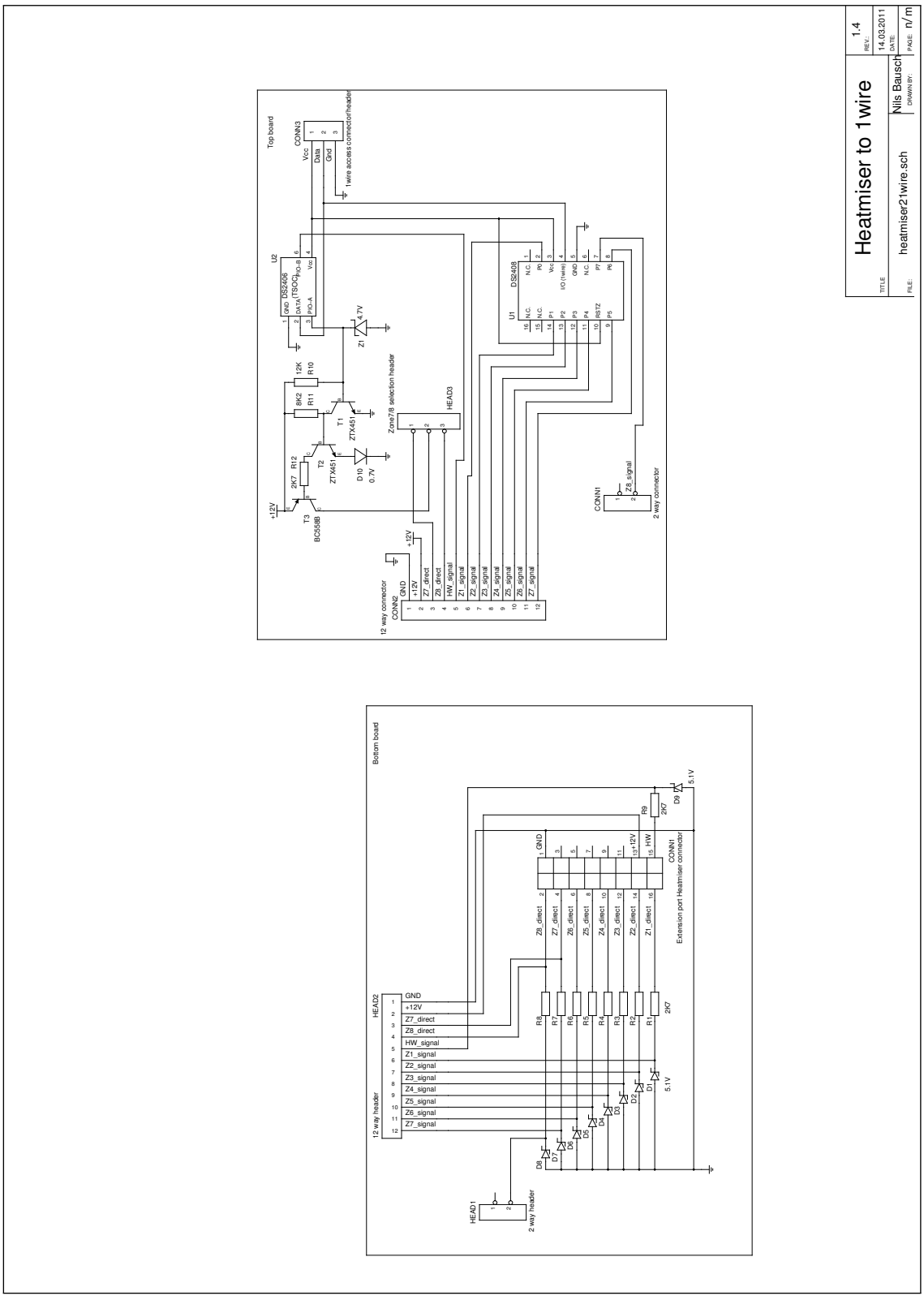


Figure B.2: *Heatmiser to 1wire interface schematic*

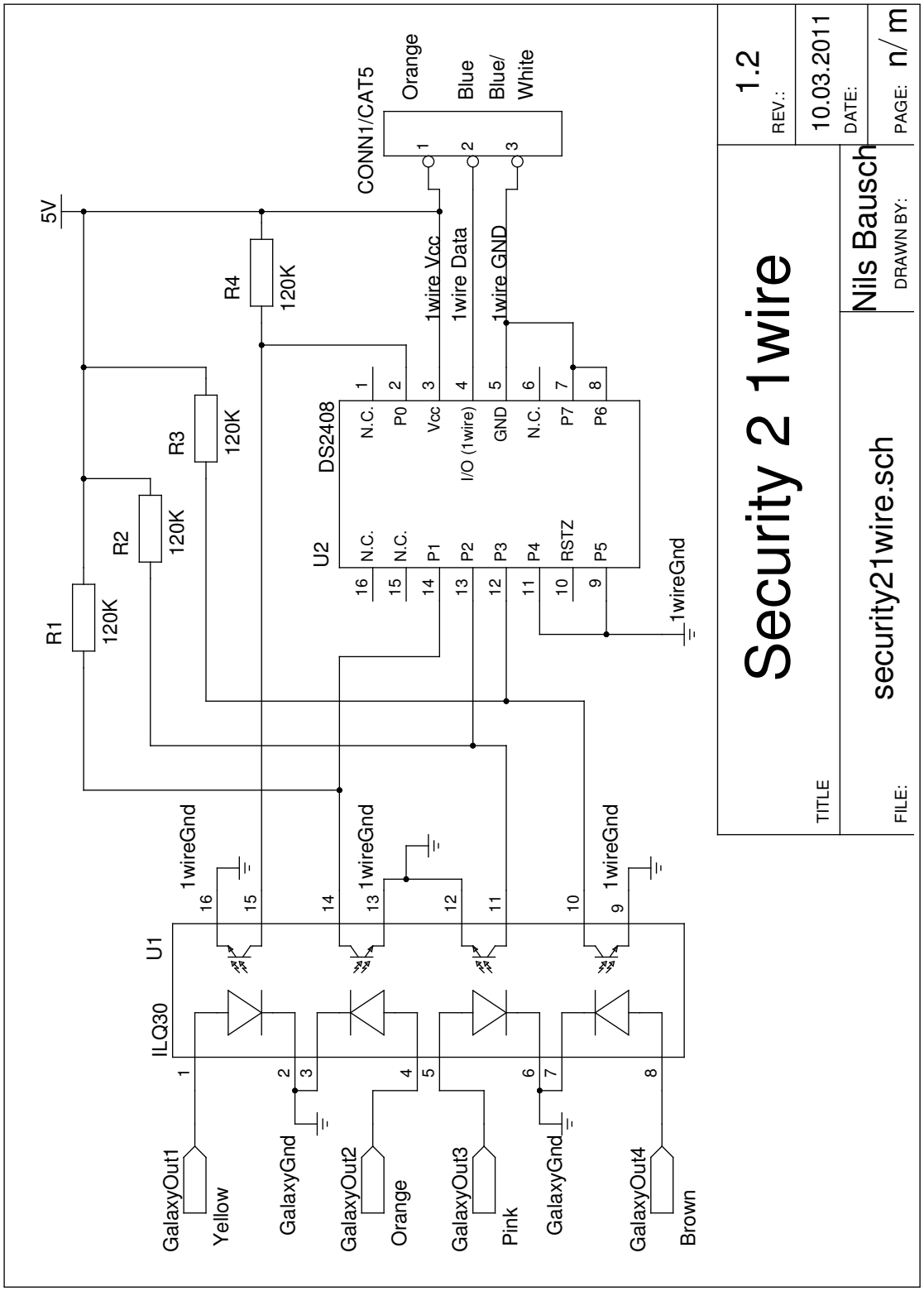


Figure B.3: *Security system to 1wire interface schematic*

Appendix C Thermal comfort

The residential home had five inhabitants and their thermal comfort was considered as input into a heating control system. With thermal comfort as a second goal, an optimized heating control was envisioned to address both energy and cost efficiency in addition to the thermal comfort of the inhabitants. This Appendix gives an overview on how thermal comfort input was collected, a short analysis on the collected data and reasoning, why it was not included in this research work.

C.1 Background

The term 'Thermal Comfort' was initially shaped by research undertaken by Fanger (1970). Thermal comfort was described as a subject's satisfaction with the thermal environment and was determined by the subject's personal opinion. The first approaches to develop an indication of thermal comfort was the Predicted Mean Vote (PMV), which was a formula incorporating several inputs to compute an index of seven comfort states. The limitations of Fanger's approach were shown in (Charles, 2003), which stated that the most fitting use cases for PMV were for Heating, Ventilation, and Air Conditioning (HVAC) controlled offices buildings with sedentary subjects. The next improvement was the introduction of Predicted Percentage Dissatisfied (PPD), which represented the percentage of dissatisfied people with certain conditions. All these approaches were taken and combined in ISO standard 7730 (BSI, 2005), which included the calculation formulas and tables for standard values.

To address residential buildings, an adaptive comfort model for naturally ventilated buildings was developed (De Dear & Brager, 1998), which described 90% and 80% acceptance boundaries of indoor temperatures for a given mean outdoor temperature. A tool developed by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) (Hoyt, Schiavon, Piccioli, Moon & Steinfeld, 2012) confirmed upon entry of parameters if conditions were within the standard ASHRAE-55 (ASHRAE, 2010). The models described by Fanger et al. and the adaptive comfort method were included in ASHRAE-55 as guidelines for engineers to

> ... specify the combinations of indoor thermal environmental factors and personal factors that will produce thermal environmental conditions acceptable to a majority of the occupants within the space.

The guidelines and standards provided a basis on which thermal comfort could be retrieved based on given input parameters.

The core temperature sensor included in the prediction models was similar to the operative temperature (BSI, 2005). The operative temperature can be defined as the average of the mean

radiant and inside air temperatures. The operative temperature was used in thermal comfort as the base temperature that was used in thermal comfort calculations and standards.

In this research it was preferred to collect thermal comfort directly from the inhabitants rather than calculated to adopt a dynamic approach. The inhabitants' inputs would have been collected over time and a database created, which would enable predictions and optimisations of thermal comfort. The aim was to create an improved adaptive thermal comfort input method for inhabitants compared to the existing standards and guidelines.

C.2 Thermal comfort input

In this Dissertation, Thermal Comfort input from the inhabitants was collected and the following states were defined:

- 'Cold' if the inhabitant felt too cold.
- 'OK' if the inhabitant felt comfortable.
- 'Hot' if the inhabitant felt too warm.

The inhabitants were asked to input this information upon return to the premises after a period of vacancy. The data would be used in a future system to turn the heating on, if it could be anticipated, that inhabitants would feel cold, which would apply air prediction models for unheated vacancy.

This approach was preferred compared with the adaptive comfort method, as the adaptive method did not account for mean outside temperatures below 10 °C, excluding colder winter season at the residential home location.

The historical data revealed, that short term vacancies exhibited temperature changes of ±0.5 °C and therefore unheated vacancies did not have an impact on thermal comfort upon return to the premises.

C.3 Collecting inhabitants' input

The user data was collected with thermal comfort input boxes, which had five user buttons and three buttons for each thermal comfort setting. The box input layout can be seen in fig. C.1. A schematic for the circuit can be seen in fig. B.1 on page 144. The circuit was built to host a eight-channel switch input sensor DS2408 and an Atmel AVR microcontroller to buffer user inputs. The thermal comfort input was read in regular intervals from the 1-wire network in the residential home and saved in the Structured Query Language (SQL) database as values '8', '16' and '24' for 'Cold', 'OK', and 'Hot' respectively.

Five boxes were created, one as a reference model and four to be used at different locations within the residential home. One box was installed on the first floor, to relate the user input to the landing air and core temperatures recorded at the same location. The remaining three comfort boxes were intended to be installed in different locations of the residential home, but never used.

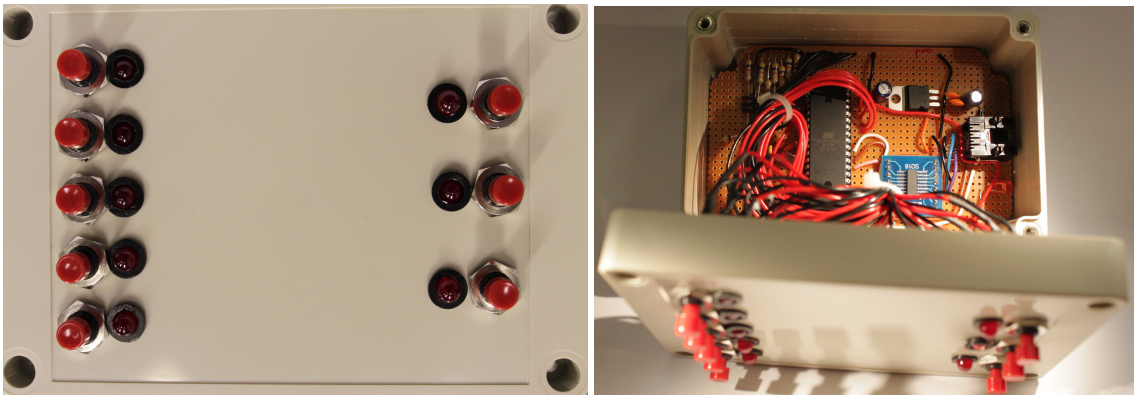


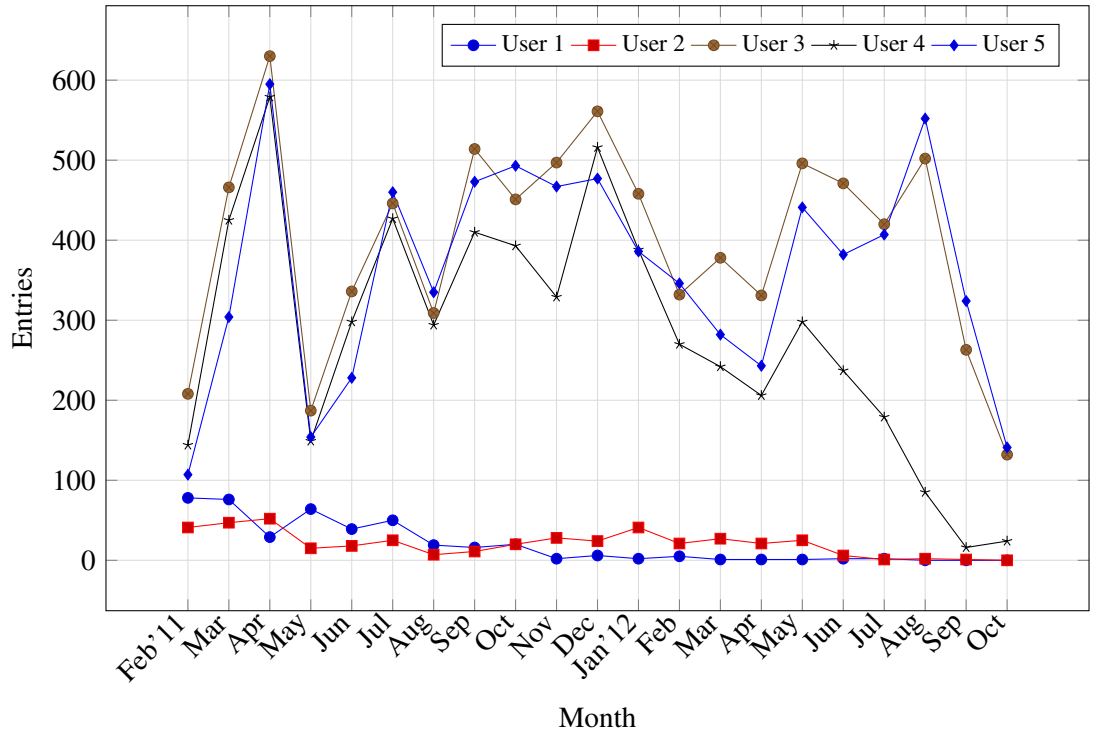
Figure C.1: *Front view with user input switches and feedback LEDs on the left and opened comfort box on the right*

C.4 Analysis of collected data

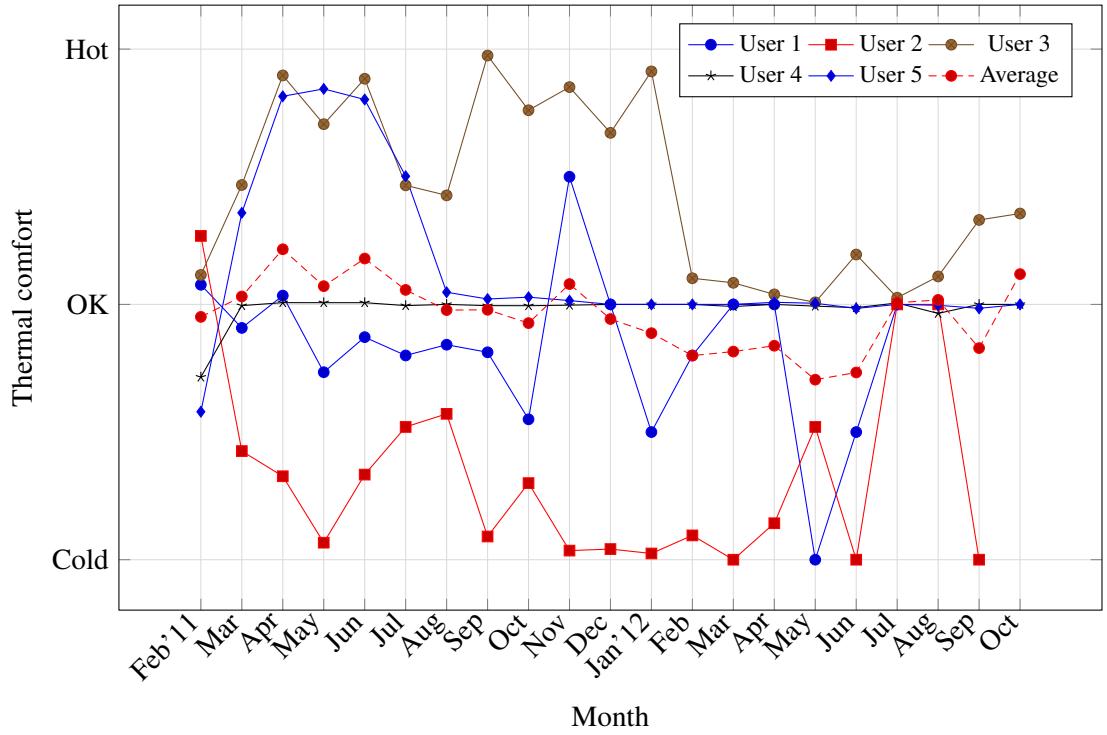
The comfort input was collected from five individuals living in the residential home. An initial analysis of the actual input can be seen in table C.1. The statistics covered a period of 606 days, from the 16th February 2011 until the 14th October 2012. The users 1 and 2 were adults and users 3, 4, and 5 were children. The averages summarized by months and users can be seen in fig. C.2b on the following page and the number of entries per user per month in fig. C.2a.

Table C.1: *Thermal comfort user input statistics*

User	24_{Hot}	16_{OK}	8_{Cold}	Total inputs
1	4.9%	85%	10.1%	982
2	8.8%	15.6%	75.5%	434
3	50%	47.8%	2.2%	8394
4	0.7%	97.7%	1.6%	5912
5	17.2%	80.5%	2.3%	7599



(a) *Thermal comfort input per user*



(b) *Average thermal comfort input per user*

Figure C.2: *Average and times of thermal comfort input per month*

C.5 Discussion

The results show, that the thermal comfort inputs were not equally spread throughout the users and users 3, 4, and 5 had a multiple of the inputs of users 1 and 2. From a discussion with the adult inhabitants, the comfort input box appeared to be a game for the children.

Furthermore, the average inputs revealed, that there was a dissent in thermal comfort opinion, for example user 3 and user 2 showed opposing preferences. The average thermal comfort of all users shown in fig. C.2b did not deviate significantly from 'OK' and was therefore not a valuable input into determining the thermal comfort of all inhabitants.

C.6 Conclusion

The thermal comfort input box was introduced to collect the inhabitant's thermal comfort. The results however showed a dissonance in the users' preferences, which made the collected data unusable in terms of finding a common thermal comfort that was desired by all inhabitants. If a common preference of thermal comfort could be identified, it could improve and add an additional goal to the optimal heating of a residential home. Improvements could also be made by enlarging the circle of respondents and residential homes, to analyse if the dissent found within the researched residential home was a general case or an outlier.

Thermal comfort was therefore not used in this research as a goal for the optimal heating.

Appendix D Ethical Review

FORM UPR16

Research Ethics Review Checklist

University of Portsmouth

Please complete and return the form to Research Section, Quality Management Division, Academic Registry, University House, with your thesis, prior to examination

Postgraduate Research Student (PGRS) Information			**Student ID:**	**413131**
Candidate Name:	**Nils Christian Bausch**			
Department:	**ENG**	**First Supervisor:**	**Dr Giles Tewkesbury**	
Start Date: (or progression date for Prof Doc students)	**01/06/2009**			

Study Mode and Route:	Part-time ■	Full-time ❑	MPhil ❑	MD ❑	PhD ■	Integrated Doctorate (NewRoute) ❑	Prof Doc (PD) ❑

Title of Thesis:	Residential home temperature prediction models that use space conditioning experiments and disparate sources of information
Thesis Word Count: (excluding ancillary data)	37493

If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study

Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).

UKRIO Finished Research Checklist:
(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: http://www.ukrio.org/what-we-do/code-of-practice-for-research/)

a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	**YES**
b) Have all contributions to knowledge been acknowledged?	**YES**
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	**YES**
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	**YES**
e) Does your research comply with all legal, ethical, and contractual requirements?	**YES**

*Delete as appropriate

UPR 16 (2013) – November 2013

Candidate Statement:	
I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)	
Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):	***535A-A942-8063-E894-0439-1274-8A34-4F1D***
Signed:  ***(Student)***	**Date: 06/02/2014**
If you have *not* submitted your work for ethical review, and/or you have answered 'No' to one or more of questions a) to e), please explain why this is so:	
Signed: ***(Student)***	**Date:**

University of Portsmouth

Certificate of Fast Track Ethics Review

Project Title:	Residential home temperature prediction models that use space conditioning experiments and disparate sources of information
Student Number:	413131
Application Date:	11/12/2013 12:33:16

You must download your referral certificate, print a copy and keep it as a record of this review.

The FEC representative for the School of Engineering is Giles Tewkesbury

It is your responsibility to follow the University Code of Practice on Ethical Standards and any Department/School or professional guidelines in the conduct of your study including relevant guidelines regarding health and safety of researchers including the following:

- University Policy
- Safety on Geological Fieldwork

It is also your responsibility to follow University guidance on Data Protection Policy:

- General guidance for all data protection issues
- University Data Protection Policy

ProjectTitle:
Residential home temperature prediction models that use space conditioning experiments and disparate sources of information

SchoolOrDepartment:
ENG

PrimaryRole:
PostgraduateStudent

SupervisorName:
Dr Giles Tewkesbury

HumanParticipants:
No

PhysicalEcologicalDamage:
No

HistoricalOrCulturalDamage:
No

HarmToAnimal:
No

Certificate Code: 535A-A942-8063-E894-0439-1274-8A34-4F1D Page 1 / 2

HarmfulToThirdParties:
No

Supervisor Review
Supervisor signature:
Date: 11/12/13